



Probabilistic Algorithms for Information and Technology Flows in the Networks

Citation

Liu, Zhenming. 2012. Probabilistic Algorithms for Information and Technology Flows in the Networks. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:9754141>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

©2012 - Zhenming Liu

All rights reserved.

Thesis advisor

Author

Michael Mitzenmacher

Zhenming Liu

Probabilistic Algorithms for Information and Technology Flows in the Networks

Abstract

This thesis studies several probabilistic algorithms for information and technology flow in the networks. Information flow refers to the circulation of information in social or communication networks for the purpose of disseminating or aggregating knowledge. Technology flow refers to the process in the network in which nodes incrementally adopt a certain type of technological product such as networking protocols. In this thesis, we study the following problems.

First, we consider the scenario where *information flow acts as media to disseminate messages*. The information flow here is considered as a mechanism of replicating a piece of information from one node to another in a network with a goal to “broadcast” the knowledge to everyone. Our studies focus on a broadcasting algorithm called the flooding algorithm. We give a tight characterization on the completion time of the flooding algorithm when we make natural stochastic assumptions on the evolution of the network.

Second, we consider the problem that *information flow acts as a device to aggregate statistics*. We interpret information flow here as artifacts produced by algorithmic procedures that serve as statistical estimators for the networks. The goal is to maintain accurate estimators with minimal information flow overhead. We study these two

problems: first, we consider the continual count tracking problem in a distributed environment where the input is an aggregate stream originating from k distinct sites and the updates are allowed to be non-monotonic. We develop an optimal algorithm in communication cost that can continually track the count for a family of stochastic streams. Second, we study the effectiveness of using random walks to estimate the statistical properties of networks. Specifically, we give the first deviation bounds for random walks over finite state Markov chains based on *mixing time properties* of the chain.

Finally, we study the problem where technology flow acts as a key to unlock innovative technology diffusion. Here, the technology flow shall be interpreted as a way to specify the circumstance, in which a node in the network will decide to adopt a new technology. Our studies focus on finding the most cost effective way to deploy networking protocols such as SecureBGP or IPv6 in the Internet. Our result is a near optimal strategy that leverages the patterns of technology flows to facilitate the new technology deployments.

Contents

Title Page	i
Abstract	iii
Table of Contents	v
Bibliographic Note	viii
Acknowledgments	ix
Dedication	xii
1 Introduction	1
1.1 Summary and outline	9
2 Information dissemination in k-dimensional spaces	11
2.1 Background	11
2.1.1 Our models and results	13
2.2 Preliminary results for random walks	16
2.3 Lower bound	33
2.3.1 Local diffusion problem	35
2.3.2 From local to global process	53
2.4 Upper bound	57
2.4.1 Leveraging local analysis	75
2.5 The case when the number of agents is sparse	90
2.6 Random walks properties	97
2.6.1 Mixing time in graphs	104
2.7 Existing techniques	105
2.7.1 Lower bound	105
2.7.2 Upper bound	108
3 Continuous distributed counting for non-monotonic streams	113
3.1 Background	113
3.1.1 Problem Definition	115
3.1.2 Our Contributions	118
3.1.3 Related Work	122

3.2	Algorithms and Notations	123
3.2.1	Building Blocks	124
3.2.2	Algorithm Overview	126
3.3	Upper Bounds	127
3.3.1	I.I.D. Input with Zero Drift	128
3.3.2	I.I.D. Input with Unknown Drift	138
3.3.3	Randomly Ordered Data Streams	151
3.3.4	Fractional Brownian Motion	162
3.4	Lower Bounds	171
3.5	Applications	180
3.5.1	Tracking the Second Frequency Moment	180
3.5.2	Bayesian Linear Regression	181
3.6	Geometric Progression Search for μ	183
4	Revisiting Chernoff-Hoeffding Bounds for Markov Chains	185
4.1	Background	185
4.2	Preliminaries	188
4.3	Chernoff-Hoeffding Bounds for Discrete Time Markov Chains	190
4.3.1	Mixing Time v.s. Spectral Expansion	196
4.3.2	Bounding the Moment Generating Function	199
4.3.3	Continuous Time Case	210
4.4	Construction of Mixing Markov Chain with No Spectral Expansion	211
4.5	The Bound When the Sum Is Less Than Mean	213
5	Technology diffusion in communication networks	218
5.1	Introduction	218
5.1.1	Our results.	220
5.2	Linearization & formulating the IP	224
5.3	Rounding algorithm.	230
5.3.1	Resolving inconsistencies using rejection-sampling	234
5.3.2	Strengthened IP and coupled sampling	239
5.4	Proof of Corollary 5.1.3	254
5.5	Lower bounds	255
5.5.1	Computational barrier	255
5.5.2	Information theoretic bound	259
5.5.3	Integrality gap	261
5.5.4	Remark on the role of flow constraints in reducing the integrality gap	275
5.6	Our problem is neither submodular nor supermodular	276
5.6.1	Fixed threshold case	277
5.6.2	Randomized threshold case	280
5.7	Experiments with the IP of Figure 5.1	284

5.8 Expository examples and figures	287
Bibliography	290
A Probability review	299
A.1 Concentration bounds	299

Bibliographic Note

All of the work in this thesis is in the process of being reviewed or published in traditional academic formats and venues. Specifically, the material in Chapter 2 appears in

Henry Lam, Zhenming Liu, Michael Mitzenmacher, Yajun Wang, Xiaorui Sun, “Information dissemination via random walks in d -dimensional spaces.”, in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2012.

The material in Chapter 3 appears in

Zhenming Liu, Božidar Radunović and Milan Vojnović, “Continuous distributed counting for non-monotonic streams”, in *ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, 2012.

The material in Chapter 4 appears in

Kai-Min Chung, Henry Lam, Zhenming Liu, Michael Mitzenmacher “Chernoff-Hoeffding bounds for Markov chains: generalized and simplified”, in *Symposium on Theoretical Aspects of Computer Science (STACS)*, 2012.

The material in Chapter 5 appears in

Sharon Goldberg, Zhenming Liu, “Technology diffusion in communication networks.”, *Boston University technical report*, Nov. 2011.

Acknowledgments

Over the years, my advisor Michael Mitzenmacher has been patiently waiting me to grow, helping me to build up my research skills, and finding me intriguing theoretical computer science questions, specifically in the areas of information theory, streaming algorithms, and stochastic processes. As a theoretician, he demonstrated to me how one can continuously extend his research horizon inside and outside theoretical computer science sheerly driven by intellectual curiosity. I could not have achieved my research goals without his help and effort.

Salil Vadhan was my advisor when I was in the Master's program at Harvard. I learned tremendous amounts about pseudorandomness and computational complexity from him, all of which has proven extremely useful throughout my time in the PhD program. I am also grateful to Leslie Valiant, Patrick Wolfe, and Yiling Chen at Harvard for serving on my qualifying exam and dissertation committee.

Kai-Min Chung was the senior graduate student I would approach when the faculty members were busy. I managed to trick him to work on several algorithmic problems outside his own research area. I also forced him to help me develop technical skills needed for the theory of computation research. I probably would not have been able to complete many of the research projects I have done without his generous help.

Outside Harvard I was fortunate to work closely with Sharon Goldberg. She gave me a very interesting algorithmic problem in communication networks when she was visiting Harvard. We worked on the problem for nearly two years and our research result was very pleasant (the result will be presented in Chapter 5 in this thesis). In this collaboration, she offered me enormous advice on tackling applied algorithm problems and making clear technical presentations.

I am also thankful to Henry Lam, who spent countless hours with me on different research projects in probability theory and optimization. He also gave me many wonderful (private) lectures in operations research and queuing theory while I was working at Boston University. I always find it very stimulating to discuss research problems with him.

I spent nearly a year as a research intern in two different labs (the theory lab in Beijing and the network and system lab in Cambridge, UK) in Microsoft Research. There, I worked with outstanding researchers over many exciting network algorithm problems. I am indebted to my supervisors at Microsoft. In particular, I worked with Wei Chen and Yajun Wang in Beijing on several network optimization problems. They shared me their priceless insight in bridging theoretical computer science with practical network problems we face in real world, which was an important experience in my growth and development. Božidar Radunović and Milan Vojnović supervised me on a distributed streaming project (our result will be presented in Chapter 3). They taught me many powerful optimization techniques that proved to be effective in our project.

I thank faculty members in the theory group of the Hong Kong University of Science and Technology (HKUST), for introducing me theoretical computer science and giving me opportunities to conduct theory research in my undergraduate days. Special thanks go to Mordecai Golin, who had been my mentor since I took his “Analysis of Algorithms” course in year one’s summer in HKUST, for offering me many independent study courses in theory, preparing me for the graduate school training, and hosting me for doing research in the summers of 2003, 2004, 2008, and

2009.

I also thank all my wonderful collaborators and coauthors, Feizhou Zhou, Xiaoming Sun, Yitong Yin, Xiaorui Sun, Hilary Finucane, Flavio Chierichetti, Vladimir Braverman, Rafail Ostrovsky, William Yuen, Christopher Thorpe, Paul Syverson, Tao Sun, Min Zhang, Chin-Yew Lin, Alex Collins, Rachel Cummings, Te Ke, David Rincon, Wei Wei, Yifei Yuan, Varun Kanade, Yanhua Li, Giorgos Zervas, Lirong Xia, Shaili Jain, and Tengyu Ma. It was amazing to work with them and hear their advice.

Working and studying in the theory lab in Maxwell Dworkin was the most memorable experience for me in graduate school. I would probably not be able to survive without these friends and colleagues, who used or frequently visited the office over the years: Adam Kirsch, Shaili Jain, Kai-Min Chung, Loizos Michael, Flavio Chierichetti, Lambert Yeung, Colin Jia Zheng, Anna Huang, Stephan Holzer, Justin Thaler, Jon Ullman, Varun Kanade, Thomas Steinke, Jiayang Jiang, James Zuo, SueYeon Chung, Scott Linderman, Alice Gao, Elaine Angelino, Elena Agapie, and Brendan Juba.

Outside Harvard, I thank Jie Wang, Xin Yuan, Kenny Huang, Pinyan Lu, Yang Yang, Yuan Zhou, Liang Li, Zeyuan Zhu for their friendships and inspiring conversations.

Finally, I would like to thank my parents Hong Liu and Yuzhong Zhou and my girlfriend Ya-Wen Shen for their unconditional love and support.

The thesis work was supported by NSF Grants CCF-0634923, CNS-0721491, and CCF-0915922.

*Dedicated to my father Hong Liu,
my mother Xiaoting Deng,
my late mother Yuzhong Zhou,
my brother Wei Liu,
and Jennifer Shen.*

Chapter 1

Introduction

This thesis studies several probabilistic algorithms for information and technology flow in large scale networks. Information flow refers to the circulation of information in social or communication networks for the purpose of disseminating or aggregating knowledge. Technology flow refers to the process in the network in which nodes incrementally adopt a certain type of technological product such as IPv6 networking protocol. While the meaning of information and technology flow can be different under different contexts, in this thesis we restrict our studies to the following scenarios.

First, we study the problem that *information flow acts as media to disseminate messages*. The information flow here is considered as a mechanism of replicating a piece of information from one node to another in a network with a goal to “broadcast” the information to everyone. Our studies focus on a simple broadcasting algorithm that is often seen in mobile ad-hoc networks (MANET), namely the *flooding algorithm*. A MANET is a self-configuring infrastructureless network of mobile devices connected

by wireless links. The availability of the communication links between devices in such network is highly volatile because the devices are expected to be moving all the time. Roughly speaking, the flooding algorithm we are interested in works as follows. In the beginning, a message needs to be broadcast to every node in the network. Upon receiving the message, each node in the network forwards the message to all of its neighbors whenever possible. This process continues until everyone becomes aware of the message. Here, the specific problem we study is to characterize the completion time of the flooding algorithm when we make natural stochastic assumptions over the evolution of the MANET.

Second, we study a pair of problems, where the *information flow acts as a device to aggregate statistics*. Here, the information flow here shall be interpreted as artifacts produced by algorithmic procedures, from which statistical estimators of the networks can be constructed. Specifically, we study the following problems with a common goal to build suitable estimators that have minimal information flow overhead. In the first problem, we study how we can continuously monitor distributed data streams in networks such as sensor networks and data centers. In a sensor network, for example, different data streams continuously arrive to different sensors. A computation task we often see is to continuously maintain statistics over the union of the data streams the sensors collect using the provably minimal amount of communication.

In the second problem, we study the effectiveness of using random walks to estimate statistical properties of the network. For example, consider a web graph $G = (V, E)$, where each $v \in V$ represents a webpage. Let S be an arbitrary subset of V . We would like to know the sum of PageRank score [90] of all the webpages

in S . A natural approach to answer this question is to carry out a (PageRank-type) random walk of length ℓ over G and use the portion of random walk that hits S as the sum estimator. Our question here is to derive a deviation bound on random walks to understand the relationship between ℓ and the accuracy of the estimator.

Finally, we study the problem where *technology flow acts as a key to unlock innovative technology diffusion*. Here, the technology flow shall be interpreted as a way to specify under what circumstance a node in the network will decide to adopt a new technology. Take the deployment of networking protocols such as SecureBGP and IPv6 as an example. From Internet Service Provider's point of view, such deployment usually comes at non-negligible cost. The new networking protocols usually will be of little value unless they are deployed to a substantial number of nodes in the networks. Thus, the Internet Service Provider will consider the adoption of the new technology only if the value of deploying the new technology is sufficiently large. It has been known that the lack of local incentive is a major obstacle for the new protocol deployment. Our question here is how we should subsidize a portion of the deployment cost so that a small set of early adopters will incentivize the deployment for all other nodes in the network.

This thesis' goal. The goal of this thesis is to develop several probabilistic and algorithmic techniques for such information and technology flow problems. The general approach we take is to first encapsulate the characteristics of the flows into a statistical or economic model and then leverage the model to design optimal algorithmic solutions. For example, to understand the performance of the flooding algorithm, we will first describe the evolution of the communication network as a natural stochastic

process. Then we will utilize the stochastic process structure to analyze the algorithm. Here, the statistical or economic models serve as a proxy to understand the average behavior of the flows in the networks, and hence usually allow us to design average case algorithms. The central contribution of this thesis is the development of quite a few techniques that fit with natural flow models.

This thesis also responds to the broader “big data” challenges we face in recent years. The past decade has witnessed extensive advances in digital sensors, communications, computation, and storage, which create huge collection of data, capturing information of value to business, science, government, and society. As suggested by Bryant et. al [17], making sense of these data require a new computation paradigm: the most appropriate strategies usually combine statistical and economic analysis, and artificial intelligence to construct models from large data sets and to infer how the system should respond to new updates. It is also argued that the key to tackling these problems boils down to building up a unified framework from computer science and statistics that gives the right tradeoff between *descriptive flexibility* and *algorithmic tractability* [81]. Descriptive flexibility refers to the expressiveness of the model to describe the data. Algorithmic tractability refers to the requirement that our solutions need to be scalable to deal with massive data sets. While unrealistic simple models with limited descriptive flexibility would be undesirable, a statistical model with good descriptive complexity could result in a solution that is not computationally trackable. Thus, the key step to deliver an effective solution is to find the right tradeoff between the model and computation limits. The results presented in this thesis emphasize the algorithmic tractability of the flow problems with the

presence of statistical or economic models. By studying fundamental flow models that exist in computer networks, we aim to explore what algorithms can and cannot do in the new computation paradigm.

Below, we walk through the problems studied in the thesis. Along the way, we will articulate the mathematical model and briefly describe the problems' motivations and our results.

Information diffusion via random walks in high dimensional space We first study the question where information flow acts as media to disseminate information. We study a natural information dissemination problem for multiple mobile agents in a bounded Euclidean space. Here, only when two agents in the network are sufficiently close, they are able to establish communication links and exchange messages. Because the nodes (agents) are expected to move all the time, the availability of the communication links between nodes is highly volatile.

We model the problem as follows. The mobile agents are placed uniformly at random in the d -dimensional space $\{-n, \dots, n\}^d$ at time zero, and one of the agents holds a piece of information to be disseminated. All the agents then perform independent random walks over the space, and the information is transmitted from one agent to another if the two agents are sufficiently close. Our goal is to bound the total time before all agents receive the information (with high probability). Our work extends Pettarin et al.'s work [92], which solved the problem for $d \leq 2$. We present tight bounds up to polylogarithmic factors for the case $d \geq 3$.

Optimal distributed counters Next, we move to study the problem of network monitoring that uses minimum communication overhead. Our studies are motivated

by the new algorithmic challenges of processing massive scale data in distributed computing environment: the current state-of-the-art amounts to daily processing of Terabytes or Petabytes of data in large data centres of major online service providers and the demand for large-scale processing seems ever increasing. It is estimated that in year 2015, the global data will reach 7,910 Exabytes of data of various types, including videos, images and data. In practice, these data are usually processed in a distributed manner via computer networks such as data centers. The major bottleneck in such systems is the communication channel, whose throughput is substantially smaller than the throughput of CPUs. Thus, it is critical to design distributed algorithms with minimum communication overhead.

In our studies, we consider the continual count tracking problem in a distributed environment where the input is an aggregate stream originating from k distinct sites and the updates are allowed to be non-monotonic, i.e. both increments and decrements are allowed. The goal is to continually track the count within a prescribed relative accuracy ϵ at the lowest possible communication cost.

Specifically, we consider an adversarial setting where the input values are selected and assigned to sites by an adversary but the order is according to a random permutation or is a random i.i.d. process. The input stream of values is allowed to be non-monotonic with an unknown drift $-1 \leq \mu \leq 1$, where the case $\mu = 1$ corresponds to the special case of a monotonic stream of only non-negative updates. We design a randomized algorithm guaranteed to track the count accurately with high probability, where the total amount of communication is sublinear to the total size of the stream and is provably optimal. Our results are substantially more positive than

the communication complexity of $\Omega(n)$ under fully adversarial input. The distributed counter can also be plugged in to other existing centralized streaming algorithms, resulting in distributed streaming algorithms with sub-linear communication cost for problems such as the tracking the second frequency moment and carrying out online Bayesian regression.

Generalized and simplified Chernoff bound for Markov chains We then continue to study the effectiveness of using random walks to estimate statistical properties of the network. Here, we study deviation bounds for random walks over possibly irreversible finite state Markov chains based on mixing time properties of the chain in both discrete and continuous time settings. Our studies confirm that using random walk is effective in aggregating network statistics even when the walk is irreversible so long as it has rapid mixing time. Such bounds would also be quite applicable to a wide range of algorithmic problems in networks. For instance, these bounds have also been applied to various online learning problem [103], testing properties of a given graph [49], analyzing the structure of the social networks [9, 85], understanding the performance of data structures [45], and computational complexity [55].

In particular, we prove the first Chernoff-Hoeffding bounds for *general (irreversible)* finite-state Markov chains based on the standard L_1 (variation distance) *mixing time* of the chain. Specifically, consider an ergodic Markov chain M and a weight function $f : [n] \rightarrow [0, 1]$ on the state space $[n]$ of M with mean $\mu \triangleq \mathbb{E}_{v \leftarrow \pi}[f(v)]$, where π is the stationary distribution of M . A t -step random walk (v_1, \dots, v_t) on M starting from the stationary distribution π has expected total weight $\mathbb{E}[X] = \mu t$, where $X \triangleq \sum_{i=1}^t f(v_i)$. Let T be the L_1 mixing time of M . We show that the probability of

X deviating from its mean by a multiplicative factor of δ , i.e. $\Pr[|X - \mu t| \geq \delta \mu t]$, is at most $\exp(-\Omega(\delta^2 \mu t/T))$ for $0 \leq \delta \leq 1$, and $\exp(-\Omega(\delta \mu t/T))$ for $\delta > 1$. The bounds hold even if the weight function f_i 's for $i \in [t]$ are distinct, provided that all of them have the same mean μ .

Technology diffusion in communication networks Finally, we move to study using cascade effects to deploy new technology over communication networks. Cascade effects provide a simple and effective way to drive global diffusion of a new technology in a network: after a few well chosen *seed* nodes are convinced to adopt the technology, more and more nodes make local decisions to adopt the technology until eventually everyone in the network has adopted it. Given the complexity and expenses involved in persuading a large, dispersed network of nodes to adopt a new technology, a particularly important algorithmic problem is to determine the *smallest* possible seedset of early adopter nodes, and thus also the “cheapest” way to drive a cascade that leads to global adoption.

Diffusion models are predicated on a model of node *utility*; namely, the benefit an individual node obtains when it decides to adopt the technology. In the rich literature on cascade effects in social networks [37,69,86,97], the model of node utility is highly local – it depends on only a node’s “friends” or immediate neighbors in social network. Here, we consider a model where a node’s utility is non-local to reflect the nature of communication network. In this case, a node’s utility depends not only on its immediate neighbors, but also on the number of (possibly distant) nodes that can communicate with using the new technology. Specifically, we consider a network $G(V, E)$ and consider a progressive technology diffusion process: a node starts out

as inactive (using an older version of the technology) and activates (adopts its new, improved version) once it obtains sufficient utility from the new technology. Once a node is active, it can never become inactive. We assume that a node u activates when its utility exceeds $\theta(u)$, i.e. if the connected component containing node u in the subgraph induced by $\{v : v \in V, \text{Node } v \text{ is active}\} \cup \{u\}$ has size at least $\theta(u)$. Our goal is to find the smallest seedset S that is feasible, i.e. when the nodes in S are activated, every other node in the graph eventually activates as well.

1.1 Summary and outline

While these four problems are studied in very different contexts, leveraging statistical or economic properties of the problems is the key ingredient in our solutions. In the information dissemination problem, our analysis critically relies on the assumptions that the agents are performing independent random walks. In the technology diffusion problem, our strategy of deploying new technology over communication network can only be implemented if we understand each autonomous agent's utility structure. In the problem of designing distributed streaming algorithms for monotonic counters, our solution relies on the input stream being randomly ordered. Finally, in our large deviation bound studies, we can only get concentration bound when walks are stochastic on graphs/networks. These extra insights on the problems' structure enable us to give solutions that are substantially better than solutions for the worst case scenarios.

We also remark that while sometimes the assumption we make may not be completely realistic, nevertheless they provide insight into what we can do in these net-

working problems, where a worst-case solution is not the primary interest.

We organize this thesis in the following way. In Chapter 2, we study the information dissemination problem. In Chapter 3, we study the problem of continuously tracking non-monotonic counter in distributed streams, and in Chapter 4, we study the large deviation bounds for finite state Markov chains. In Chapter 5, we study the technology diffusion problem.

Chapter 2

Information dissemination in k -dimensional spaces

2.1 Background

We study the following information diffusion problem: let a_1, a_2, \dots, a_m be m agents initially starting at locations chosen uniformly at random in $\mathcal{V}^d = \{-n, -(n-1), \dots, n\}^d$ and performing independent random walks over this space. One of the agents initially has a message, and the message is transmitted from one agent to another when they are sufficiently close. We are interested in the time needed to flood the message, that is, the time when all agents obtain the message. In other settings, this problem has been described as a virus diffusion problem, where the message is replaced by a virus that spreads according to proximity. We use *information diffusion* and *virus spreading* interchangeably, depending on which is more useful in context. This is a natural model that has been extensively studied. For example, Alves et al.

and Kesten et al. coined the name “frog model” for this problem in the virus setting, and studied the shape formed by the infected contour in the limiting case [4, 5, 71]. In the flooding time setting, early works used a heuristic approximation based on simplifying assumptions to characterize the dynamics of the spread of the message [7, 72, 107]. More recent works provide fully rigorous treatments under this or similar random walk models [25, 26, 91, 92, 101].

The most relevant recent works are those of Pettarin et al. [92] and Peres et al. [91, 101]. The work of Pettarin et al. examines the same model as ours, but their analysis is only for one- and two-dimensional grids. The work of Sinclair and Stauffer [101] considers a similar model they call mobile geometric graphs, and their work extends to higher dimensions. However, their focus and model both have strong differences from ours. For example, they assume a Poisson point process of constant intensity, leading to a number of agents linear in the size of the space. In contrast, our results allow a sublinear number of agents, a scenario not directly relevant to their model. Also, they focus on structural aspects on the mobile graphs, such as percolation, while we are primarily interested in the diffusion time. There are additional smaller differences, but the main point is that for our problem we require and introduce new techniques and analysis.

This chapter presents matching lower bounds and upper bounds (up to polylogarithmic factors) for the flooding problems in d -dimensional space for an arbitrary constant d . For ease of exposition, in this chapter we focus on the specific case where $d = 3$, which provides the main ideas. Two- and three- dimensional random walks have quite different behaviors – specifically, two-dimensional random walks are recur-

rent while three-dimensional random walks are transient – so it is not surprising that previous results for two dimensions fail to generalize immediately to three-dimensional space. Our technical contributions include new techniques and tools for tackling the flooding problem by building sharper approximations on the effect of agent interactions.

Although the information diffusion problem in three or more dimensions appears less practically relevant than the two-dimensional case, we expect the model will still prove valuable. For instance, particles in a high dimensional space may provide a latent-space representation of the agents in a dynamic social network [58, 96], so understanding information diffusion process may be helpful for designing appropriate latent space models in the future. Also, the problem is mathematically interesting in its own right.

2.1.1 Our models and results

We follow the model developed in [92]. Let $\mathcal{V}^d = \{-n, -(n-1), \dots, 0, \dots, (n-1), n\}^d$ be a d -dimensional grid. Let $A = \{a_1, a_2, \dots, a_m\}$ be a set of moving agents on \mathcal{V}^d . At $t = 0$, the agents spread over the space according to some distribution \mathcal{D} . Throughout this chapter, we focus on the case where \mathcal{D} is uniform. Agents move in discrete time steps. Every agent performs a symmetric random walk defined in the natural way. Specifically, at each time step an agent not at a boundary moves to one of its $2d$ neighbors, each with probability $1/(2d)$. If an agent is at a boundary, so there is no edge in one or more directions, we treat each missing edge as a self-loop. Let $\Xi_1(t), \dots, \Xi_m(t) \in \{0, 1\}$ each be a random variable, where $\Xi_i(t)$ represents whether

the agent a_i is infected at time step t . We assume $\Xi_1(0) = 1$ and $\Xi_i(0) = 0$ for all $i \neq 1$. The value $\Xi_i(t)$ will change from 0 to 1 if at time t it is within distance 1 to another infected agent a_j . (We use distance 1 instead of distance 0 to avoid parity issues.) Once a value $\Xi_j(t)$ becomes 1, it stays 1.

Definition 2.1.1. (Information diffusion problem). *Let $A_1, A_2, \dots, A_m \in \mathcal{V}^d$ be the initial positions of the agents a_1, \dots, a_m and let $S_t^1(A_1), S_t^2(A_2), \dots, S_t^m(A_m)$ be m independent random walks starting at A_1, \dots, A_m respectively, so that $S_t^i(P)$ is the position of agent a_i at time t given that at $t = 0$ its position was $P \in \mathcal{V}^d$. The infectious state of each agent at time step t is a binary random variable $\Xi_i(t)$ such that*

- $\Xi_1(0) = 1, \Xi_i(0) = 0$ for all other i , and
- for all $t > 0$, $\Xi_i(t) = 1$ if and only if

$$(\Xi_i(t-1) = 1) \quad \text{or} \quad (\exists j : \Xi_j(t-1) = 1 \wedge \|S_t^i(A_i) - S_t^j(A_j)\|_1 \leq 1).$$

We define the finishing time of the diffusion process, or the diffusion time, as $T = \inf\{t \geq 0 : |\{\Xi_i(t) = 1\}| = m\}$.

The following results for the diffusion time for 1 and 2 dimensional spaces are proved in [92].

Theorem 2.1.2. *Consider the information diffusion problem for $d = 1, 2$ dimensions, and assume the agents are initially uniformly distributed over \mathcal{V}^d . Then, with high probability,*

$$T = \tilde{\Theta}(n^2 \cdot m^{-1/d}). \tag{2.1}$$

It is natural to ask whether Equation 2.1 also holds for $d \geq 3$. Our results show this is not the case.

Theorem 2.1.3. (Diffusion time for $d \geq 3$) *Consider the information diffusion problem for $d \geq 3$ with initially uniformly distributed agents over \mathcal{V}^d . Then there exists a constant c such that*

$$\begin{aligned} \text{if } cn^{d-2} \log^2 n < m < n^d : \quad T &= \tilde{\Theta}(n^{d/2+1} \cdot m^{-1/2}) \text{ with high probability;} \\ \text{if } m < cn^{d-2} \log^{-2} n : \quad T &\leq \tilde{\Theta}(n^d/m) \text{ with high prob.} \\ \text{and } T &\geq \tilde{\Theta}(n^d/m) \text{ almost surely.} \end{aligned} \tag{2.2}$$

Notice that Theorems 2.1.3 and 2.1.2 yield the same result for $d = 2$, as well as when $d = 1$ and $m = \Theta(n)$. Here when we say with high probability, we mean the statement holds with probability $1 - n^{-\gamma}$ for any constant γ and suitably large n . When we say almost surely, we mean with probability $1 - o(1)$. When $m \geq n^d$, the result is implicit in [71] and the diffusion time in this case is $\tilde{\Theta}(n)$. Finally, there are some technical challenges regarding the case $cn^{d-2} \log^{-2} n \leq m \leq cn^{d-2} \log^2 n$ that we expect to address in the future.

An interesting point of our result is that when the number of agents m is greater than n^{d-2} , the finishing time is less than the mixing time of each individual random walk, and therefore the analysis requires techniques that do not directly utilize the mixing time. The rest of this chapter focuses on deriving both the lower and upper bounds for this interesting case; the case where $m < cn^{d-2} \log^{-2} n$, which harnesses similar ideas and a mixing time argument, is only briefly described at the end. Finally, as previously mentioned, we provide only the analysis for the three dimensional case, and note that the results can be generalized to higher dimensions.

Theorem 2.1.3 can also be expressed in the terms of the density of agents. Let $\lambda = m/n^d$ be the density. We can express the diffusion time as $T = \tilde{\Theta}(n/\sqrt{\lambda})$ w.h.p. for $cn^{-2}\log^2 n < \lambda < 1$, whereas for $\lambda < cn^{-2}\log^{-2} n$ we have $T \leq \tilde{\Theta}(1/\lambda)$ w.h.p. and $T \geq \tilde{\Theta}(1/\lambda)$ almost surely.

We remark that all theorems/propositions/lemmas in this chapter are assumed to hold for sufficiently large n , but for conciseness we may not restate this condition in every instance.

2.2 Preliminary results for random walks

In this section we lay out some preliminary results on random walks that will be useful in the subsequent analysis (For completeness, Chapter 2.6 also presents other known properties of random walks that we will use throughout this Chapter). These results focus on probabilistic estimates for the meeting time/position of multiple random walks. Along the way, we will also illustrate the limitations of some of these estimates, hence leading to the need of more sophisticated techniques in our subsequent analysis.

Let \mathbf{Z} be the set of integers, and \mathbf{Z}^3 be the set of integral lattice points in \mathbf{R}^3 . For two points $A, B \in \mathbf{Z}^3$, we write $A - B$ as the 3-dimensional vector pointing from B to A . For a vector $\vec{x} \in \mathbf{R}^3$, denote the i th coordinate of \vec{x} as x_i . Define the L_p norm of a vector as $\|\vec{x}\|_p = (\sum_{i \leq 3} |x_i|^p)^{1/p}$, and also the infinite-norm in the standard manner $\|\vec{x}\|_\infty = \max_{i \leq 3} |x_i|$. We moderately overload x in this chapter, i.e. x is a scalar and \vec{x} is a vector.

Let S^1 and S^2 be two random walks in either \mathcal{V}^3 (bounded walks) or \mathbf{Z}^3 (un-

bounded walks). We say two walks S^1 and S^2 *meet* at time t if their L_1 -distance is within 1 at that time and two walks S^1 and S^2 *collide* at t if they are exactly at the same position at time t .

Definition 2.2.1 (Passage probability). *Let S be a random walk in \mathbf{Z}^3 starting at the origin O . Let B be a point with $B - O = \vec{x}$ (which is a three dimensional vector). Define the probability that S is at B at time t as $p(t, \vec{x})$. Define the probability that S visits B within time t as $q(t, \vec{x})$.*

We want to characterize the chance that two or more random walks in either \mathcal{V}^3 or \mathbf{Z}^3 meet. More specifically, consider the following question. Let A_1, \dots, A_j , and B be $j + 1$ points over the 3-dimensional space \mathbf{Z}^3 such that for all $i \in [j]$, the L_1 distance between A_i and B is $\|A_i - B\|_1 \geq x$. Let $S^1(A_1), \dots, S^j(A_j), S^{j+1}(B)$ be independent random walks that start with these points respectively. Our goal is to understand the probability that all the walks S^1, \dots, S^j will meet or collide with the random walk S^{j+1} within x^2 time steps. We note that if the agents starting at B was stationary instead of following its own random walk then the analysis of the situation would be straightforward. In particular, the probability that all the walks would intersect B is $\tilde{O}(1/x^k)$. This follows from standard results, including Theorem 2.6.6 and Lemma 2.6.7 provided in Chapter 2.6. We need to consider a more challenging situation when the agent starting at B is also moving.

To begin, we shall consider the case where $j = 1$, so that we have just two moving agents.

Definition 2.2.2. *Let A and B be two points over \mathbf{Z}^3 such that $A - B = \vec{x}$. where $\|\vec{x}\|_1$ is an even number. Let S^1 and S^2 be two independent unbounded random walks*

that start at A and B respectively. Define $Q(t, \vec{x})$ as the probability that S^1 and S^2 collide before time t .

We can use a simple coupling argument to relate $Q(t, \vec{x})$ with $q(t, \vec{x})$. The result is described as follows.

Lemma 2.2.3. *Let A and B be two points over \mathbf{Z}^3 such that $A - B = \vec{x}$, where $\|\vec{x}\|_1$ is an even number. Consider $Q(t, \vec{x})$ and $q(t, \vec{x})$ defined above. We have $Q(t, \vec{x}) = q(2t, \vec{x})$. Furthermore, for $t \geq \|\vec{x}\|_2^2$, $Q(t, \vec{x}) = \Theta(1/\|\vec{x}\|_2)$.*

Proof of Lemma 2.2.3. Let us consider the following two processes \mathbb{P}_1 and \mathbb{P}_2 in the same probability space (we slightly abuse the terminology “process” to mean the expression of the random tosses that drive all random walks of interest).

1. The process \mathbb{P}_1 : consider the random walk $S(A)$. We are interested in the event that $S_t(A)$ visits B within time $2t$, which occurs with probability $q(2t, \vec{x})$. Notice that $S(A)$ is unable to visit B at odd steps.
2. The process \mathbb{P}_2 : consider the random walks $S^1(A)$ and $S^2(B)$. We are interested in the event that the two walks collide by the time t , which occurs with probability $Q(t, \vec{x})$.

We couple the two random processes as follows. We first construct the single random walk in \mathbb{P}_1 from the two walks in \mathbb{P}_2 . Note that one time step in \mathbb{P}_2 involves simultaneous moves of the walks $S^1(A)$ and $S^2(B)$. Corresponding to this step, the single walk in \mathbb{P}_1 will be set to move first in the same direction as $S^1(A)$, and then in the reverse direction from $S^2(B)$. This way the moves at time $t > 0$ in \mathbb{P}_2 are translated into the moves at time $2t - 1$ and $2t$ in \mathbb{P}_1 . The construction can naturally

be reversed to map a walk in \mathbb{P}_1 to two walks in \mathbb{P}_2 . This coupling ensures the L_1 distance between S^1 and S^2 at time t in \mathbb{P}_2 is the same as the distance between S and B at time $2t$ in \mathbb{P}_1 . Note that collision in \mathbb{P}_2 can only occur at even steps, and hence the hitting event in \mathbb{P}_1 is well-defined. Therefore S^1 and S^2 collide at or before t if and only if S visits B at or before $2t$.

Using the bound given in Lemma 2.6.7, we have for $t \geq \|\vec{x}\|_2^2$,

$$Q(t, \vec{x}) = \Theta\left(\frac{1}{\|\vec{x}\|_2}\right)$$

□

Next, let us move to the case of j random walks in \mathbf{Z}^3 , in which $j > 1$.

Lemma 2.2.4. *Let A_1, A_2, \dots, A_j , and B be points in \mathbf{Z}^3 such that $\|A_i - B\|_1$ are even and $\|A_i - B\|_1 \geq x$ for all $i \leq j$. Let $S^1(A_1), \dots, S^j(A_j), S^{j+1}(B)$ be $j+1$ independent random walks that start at A_1, \dots, A_j, B respectively. Let $t = x^2$. Then the probability that all the walks S^1, \dots, S^j collide with S^{j+1} within time t is at most $\left(\frac{\zeta j}{x}\right)^j$, where ζ is a sufficiently large constant.*

Proof of Lemma 2.2.4. Let $\Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j}$ be the event that S^i and S^{j+1} collide at C_i at time step t_i (not necessarily for the first time) for all $i \in [j]$. Our goal is to bound the following quantity

$$\begin{aligned} & \Pr \left[\exists t_1, \dots, t_j, C_1, \dots, C_j : \Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1 \right] \\ & \leq j! \Pr \left[\exists t_1 \leq \dots \leq t_j, C_1, \dots, C_j : \Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1 \right] \\ & \leq j! \sum_{t_1 \leq t_2 \leq \dots \leq t_j} \sum_{C_1, \dots, C_j} \Pr[\Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1]. \end{aligned}$$

We cut the time interval into j frames $[0, t_1]$, $[t_1, t_2]$, ..., $[t_{j-1}, t_j]$, so that the random walks with different frames are independent. Define D_{i-1} be the position of S^i at time t_{i-1} . For notational convenience, we let $D_0 = A_1$, $C_0 = B$, and $t_0 = 0$.

The event $\Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1$ implies that in the i -th time interval $[t_{i-1}, t_i]$ we have

1. S^i moves from A_i at time 0 to D_{i-1} at time t_{i-1} .
2. at time t_{i-1} , the walk S^{j+1} is at C_{i-1} .
3. at time t_i , the walk S^{j+1} and the walk S^i are both at C_i .

By standard results regarding high dimensional random walks (e.g. see Theorem 2.6.5), the probability that the first event happens is at most

$$\frac{3}{t_{i-1}^{1.5}} \left(\frac{1.5}{2\pi} \right)^{1.5} \exp \left\{ \frac{-3\|A_i - D_{i-1}\|_2^2}{2t_{i-1}} \right\} \quad (2.3)$$

and the probability that both the second and the third events happen is at most

$$\frac{3}{(t_i - t_{i-1})^3} \left(\frac{1.5}{2\pi} \right)^{1.5} \exp \left\{ \frac{-3(\|D_{i-1} - C_i\|_2^2 + \|C_{i-1} - C_i\|_2^2)}{2(t_i - t_{i-1})} \right\}. \quad (2.4)$$

The error term in Theorem 2.6.5 is swallowed by the larger leading constants 3 in Equations 2.3 and 2.4.

As S^i 's walk before time t_{i-1} is independent to the walks of S^i and S^{j+1} between t_{i-1} and t_i , the probability that the three subevents above happen can be bounded by taking the product of Equation 2.3 and 2.4 above.

Let

$$f_i = \frac{3}{(t_i - t_{i-1})^3} \left(\frac{1.5}{2\pi} \right)^{1.5} \exp \left\{ \frac{-3(\|D_{i-1} - C_i\|_2^2 + \|C_{i-1} - C_i\|_2^2)}{2(t_i - t_{i-1})} \right\} \text{ for } 1 \leq i \leq j, \quad (2.5)$$

$$g_i = \frac{3}{t_{i-1}^{1.5}} \left(\frac{1.5}{2\pi} \right)^{1.5} \exp \left\{ \frac{-3\|A_i - D_{i-1}\|_2^2}{2t_{i-1}} \right\} \text{ for } 2 \leq i \leq j, \quad (2.6)$$

We also let $g_1 = 1$ and $g = \prod_{i \leq j} g_i$. We have

$$\Pr[\exists t_1 \leq \dots \leq t_j, C_1, \dots, C_j : \Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1] \leq \sum_{t_1, \dots, t_j} \sum_{C_1, \dots, C_j} \sum_{D_1, \dots, D_{j-1}} (f_1 g_1)(f_2 g_2) \dots (f_j g_j)$$

We now carefully bound this sum. Observe that in Equation 2.5, when $t_i - t_{i-1}$ is fixed and $\|D_{i-1} - C_i\|_2^2 + \|C_{i-1} - C_i\|_2^2$ is sufficiently large, the quantity f_i asymptotically becomes

$$\exp(-\Theta(\max\{\|D_{i-1} - C_i\|_2^2, \|C_{i-1} - C_i\|_2^2\})).$$

This motivates us to group the triples $\{C_{i-1}, C_i, D_{i-1}\}$ together, where the triples are covered by balls with approximately the same size under the L_∞ norm. Specifically, we let \mathbb{D}_r be the set of triples (A, B, C) where $A, B, C \in \mathbf{Z}^3$ and $\max\{\|A - B\|_1, \|A - C\|_1, \|B - C\|_1\} \leq r$. Also, we say $\{A, B, C\} \in \partial \mathbb{D}_r$ if $\{A, B, C\} \in \mathbb{D}_r - \mathbb{D}_{r-1}$. Notice by telescoping, we have $\mathbb{D}_r = \bigcup_{i \leq r} \partial \mathbb{D}_i$. We may thus group the variables C_i and D_i by parameterizing the radii of the balls,

$$\begin{aligned} & \Pr[\exists t_1 \leq \dots \leq t_j, C_1, \dots, C_j : \Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1] \\ & \leq \sum_{t_1, \dots, t_j} \sum_{C_1, \dots, C_j} \sum_{D_1, \dots, D_{j-1}} (f_1 g_1)(f_2 g_2) \dots (f_j g_j) \\ & = \sum_{C_1 \in \mathcal{V}^3} \sum_{r_1 \geq 0} \sum_{r_2 \geq 0} \dots \sum_{r_{j-1} \geq 0} \sum_{\substack{\{C_1, C_2, D_1\} \\ \in \partial \mathbb{D}_{r_1}}} \sum_{\substack{\{C_3, D_2\} \\ \{C_2, C_3, D_2\} \\ \in \partial \mathbb{D}_{r_2}}} \dots \sum_{\substack{\{C_j, D_{j-1}\} \\ \{C_{j-1}, C_j, D_{j-1}\} \\ \in \partial \mathbb{D}_{r_{j-1}}}} \sum_{t_1 < \dots < t_j} f_1 \cdot f_2 \cdot \dots \cdot f_j \cdot g \end{aligned}$$

First observe that by the triangle inequality $\|A - C_1\|_1 + \|B - C_1\|_1 \geq \|A - B\|_1 = x$, and for any vector $\vec{v} \in \mathbf{R}^3$,

$$\frac{1}{\sqrt{3}} \|\vec{v}\|_1 \leq \|\vec{v}\|_2 \leq \|\vec{v}\|_1. \quad (2.7)$$

We have

$$\|D_0 - C_1\|_2^2 + \|C_0 - C_1\|_2^2 = \|A - C_1\|_2^2 + \|B - C_1\|_2^2 \geq \frac{1}{3}(\|A - C_1\|_1^2 + \|B - C_1\|_1^2) \geq \frac{x^2}{6}. \quad (2.8)$$

Next, by the triangle inequality again, $\|D_i - C_{i+1}\|_1 + \|C_i - C_{i+1}\|_1 \geq \|D_i - C_i\|_1$.

Meanwhile, we have

$$\max\{\|D_i - C_{i+1}\|_1, \|C_i - C_{i+1}\|_1, \|D_i - C_i\|_1\} = r_i.$$

Together with the relationship between the L_1 and L_2 norms in Equation 2.7, we obtain

$$\|D_i - C_{i+1}\|_2^2 + \|C_i - C_{i+1}\|_2^2 \geq r_i^2/6 \quad \text{for } 1 \leq i < j.$$

Next, for $i \geq 2$ we define

$$\hat{f}_i = \frac{1}{(t_i - t_{i-1})^3} \exp \left\{ \frac{-r_{i-1}^2}{4(t_i - t_{i-1})} \right\},$$

and define

$$\hat{g} = \frac{1}{(t_1 \dots t_{j-1})^{1.5}}.$$

It is clear that $f_i \leq \hat{f}_i$ for all $i \geq 2$ and $g \leq \hat{g}$. For notational convenience, we let $\hat{f}_1 = f_1$.

Our goal is now to bound the term

$$\begin{aligned} \eta &\equiv \sum_{r_1, \dots, r_{j-1}} \sum_{\substack{\text{all triples} \\ \{C_i, C_{i+1}, D_i\}}} \sum_{t_1 \leq \dots \leq t_j} \left(\prod_{i \leq j} \hat{f}_i \right) \cdot \hat{g} \\ &= \sum_{r_1, \dots, r_{j-1}} \sum_{\substack{\text{all } i < j: \\ \{C_i, C_{i+1}, D_i\}}} \sum_{t_1} \hat{f}_1 \underbrace{\frac{1}{t_1^{1.5}}}_{\text{from } \hat{g}} \sum_{t_2} \hat{f}_2 \underbrace{\frac{1}{t_2^{1.5}}}_{\text{from } \hat{g}} \dots \sum_{t_{j-1}} \hat{f}_{j-1} \underbrace{\frac{1}{t_{j-1}^{1.5}}}_{\text{from } \hat{g}} \sum_{t_j} \hat{f}_j. \end{aligned}$$

Next, let us rearrange the indices and decompose the quantity into different parts (in terms of Υ_i defined below) and express η as in (2.9).

$$\underbrace{\sum_{t_1 \geq 0} \sum_{C_1 \in \mathcal{V}^3} \frac{\hat{f}_1}{t_1^{1.5}} \left(\sum_{\substack{r_1 \geq 0 \\ t_2 \geq t_1}} \sum_{\substack{C_2, D_1 \\ \{C_1, C_2, D_1\} \in \partial \mathbb{D}_{r_1}}} \frac{\hat{f}_2}{t_2^{1.5}} \left(\sum_{\substack{r_2 \geq 0 \\ t_3 \geq t_2}} \sum_{\substack{C_3, D_2 \\ \{C_2, C_3, D_2\} \in \partial \mathbb{D}_{r_2}}} \frac{\hat{f}_3}{t_3^{1.5}} \dots \left(\sum_{\substack{r_{j-2} \geq 0 \\ t_{j-1} \geq t_{j-2}}} \sum_{\substack{C_{j-1}, D_{j-2}: \\ \{C_{j-1}, C_{j-2}, D_{j-2}\} \in \partial \mathbb{D}_{r_{j-2}}} \underbrace{\left(\sum_{\substack{r_{j-1} \geq 0 \\ t_j \geq t_{j-1}}} \sum_{\substack{C_j, D_{j-1}: \\ \{C_j, C_{j-1}, D_{j-1}\} \in \partial \mathbb{D}_{r_{j-1}}} \hat{f}_j \right)}_{\Upsilon_j} \right) \right)_{\Upsilon_{j-1}} \right)_{\Upsilon_3} \right)_{\Upsilon_2} \right)_{\Upsilon_1}}_{(2.9)}$$

Let us briefly interpret the meaning of Υ_i : this term describes an upper bound for the following two groups of events:

- the collisions between S^{j+1} and S^i, S^{i+1}, \dots , and S^j at time t_i, t_{i+1}, \dots, t_j respectively.
- the fact that at time $t_{i'}$ the walk $S^{i'+1}$ is at $D_{i'}$ for all $i \leq i' \leq j$ (i.e., $S_{t_{i'}}^{i'+1} = D_{i'}$).

which is conditioned on knowing the values for

$$\mathfrak{S} = \{t_1, \dots, t_{i-1}, r_1, \dots, r_{i-1}, C_1, \dots, C_{i-1}, D_1, \dots, D_{i-2}\}.$$

When C_{i-1} is known, this information imposes a constraint over the way to enumerate C_i and D_{i-1} because we require $\{C_{i-1}, C_i, D_{i-1}\} \in \partial\mathbb{D}_{r_{i-1}}$ for a specific r_{i-1} . Therefore, the computation of Υ_i depends on the value of C_{i-1} . A second constraint imposed from knowing \mathfrak{S} is that we need $t_{i-1} \leq t_i \leq \dots \leq t_j$. Υ_i does not depend on other values in \mathfrak{S} . In what follows, we write Υ_i as a function of C_{i-1} and t_{i-1} .

Specifically, let us define the function Υ_i in a forward recursive manner (the summations of r_i and t_i are over integers):

$$\Upsilon_i = \begin{cases} \sum_{r_{j-1} \geq 0} \sum_{\substack{C_j, D_{j-1}: \\ \{C_{j-1}, C_j, D_{j-1}\} \\ \in \partial\mathbb{D}_{r_{j-1}}}} \sum_{t_{j-1} \leq t_j \leq x^2} \hat{f}_j & \text{if } i = j \text{ (base case)} \\ \sum_{r_{i-1} \geq 0} \sum_{\substack{C_i, D_{i-1}: \\ \{C_{i-1}, C_i, D_{i-1}\} \\ \in \partial\mathbb{D}_{r_{i-1}}}} \sum_{t_i \geq t_{i-1}} \left(\hat{f}_i \frac{1}{t_i^{1.5}} \Upsilon_{i+1} \right) & \text{if } 1 < i < j \\ \sum_{t_1, C_1} \frac{\hat{f}_1}{t_1^{1.5}} \Upsilon_2 & \text{if } i = 1. \end{cases} \quad (2.10)$$

The variable Υ_1 is the quantity we desire to bound. Let $\Delta t_i = t_i - t_{i-1}$ for all i (and we shall let $t_0 = 0$). Let us start with bounding

$$\Upsilon_j = \sum_{r_{j-1}} \sum_{C_j, D_{j-1}} \sum_{\Delta t_j > 0} \frac{3}{\Delta t_j^3} \exp\left(\frac{-r_{j-1}^2}{4\Delta t_j}\right)$$

We shall first find the total number of $\{C_j, D_{j-1}\}$ pairs so that $\{C_j, C_{j-1}, D_{j-1}\} \in \partial\mathbb{D}_{r_{j-1}}$. Notice that when r_{j-1} and C_{j-1} are fixed, at least one of $\|C_{j-1} - C_j\|_1$, $\|C_{j-1} - D_{j-1}\|_1$, and $\|C_j - D_{j-1}\|_1$ is exactly r_{j-1} . When $\|C_{j-1} - D_{j-1}\|_1 = r_{j-1}$, the number of possible D_{j-1} is $4r_{j-1}(r_{j-1} - 1) \leq 4r_{j-1}^2$. An upper bound on the number of possible C_j is $4r_{j-1}^3$. Therefore, when $\|C_{j-1} - D_{j-1}\|_1 = r_{j-1}$, the number of $\{C_j, D_{j-1}\}$ pairs is at most $16r_{j-1}^5$. We may similarly analyze the other two cases to find that the total number of $\{C_j, D_{j-1}\}$ pairs is at most $48r_{j-1}^5$. Thus, we have

$$\begin{aligned}
\Upsilon_j &= \sum_{r_{j-1}} \sum_{C_j, D_{j-1}} \sum_{\Delta t_j^3} \frac{3}{\Delta t_j^3} \exp\left(\frac{-r_{j-1}^2}{4\Delta t_j}\right) \\
&= \sum_{\Delta t_j} \frac{1}{\Delta t_j^3} \left(\sum_{r_{j-1}} 3 \times 48r_{j-1}^5 \exp\left(\frac{-r_{j-1}^2}{4\Delta t_j}\right) \right) \\
&\leq \sum_{\Delta t_j} \frac{1}{\Delta t_j^3} \left(2 \cdot \int_0^\infty 144r_{j-1}^5 \exp\left(\frac{-r_{j-1}^2}{4\Delta t_j}\right) dr_{j-1} \right) \\
&= \sum_{\Delta t_j} \frac{1}{\Delta t_j^3} 18432 \Delta t_j^3 \\
&= 18432x^2 \leq \zeta_0 x^2,
\end{aligned}$$

where $\zeta_0 = 18432$. The last equality holds because we are considering a time frame of length x^2 and therefore $\Delta t_j \leq x^2$. Let us explain the derivation in greater detail because similar techniques will be used again in the rest of the analysis. Define $h(x) = x^5 \exp\left(-\frac{x^2}{4\Delta t_j}\right)$. The function $h(x)$ is a unimodal function with a unique

global maximal value. Let $x_0 = \arg \inf_{x \geq 0} h(x)$. Then we have

$$\begin{aligned} \sum_{x \in \mathbf{N}} h(x) &\leq \sum_{x=1}^{\lfloor x_0 \rfloor} h(x) + \sum_{x=\lceil x_0 \rceil}^{+\infty} h(x) \\ &\leq \int_1^{x_0+1} h(x) dx + \int_{\lfloor x_0 \rfloor}^{\infty} h(x) dx \\ &\leq 2 \int_0^{\infty} h(x) dx. \end{aligned}$$

While this bound is quite rough, it suffices for our purpose; the same approach is used to bound the summation of unimodal functions elsewhere. The third equality holds because of the following fact,

$$\int_0^{\infty} x^5 \exp\left(-\frac{x^2}{4\ell}\right) dx = 64\ell^3 \quad (2.11)$$

for any ℓ . (This can be verified through standard software packages such as Mathematica).

We can prove the following hypothesis for Υ_i :

$$\text{for all } 1 \leq \ell \leq j-2 : \Upsilon_{j-\ell}(C_{j-\ell-1}, t_{j-\ell-1}) \leq x^2 \zeta_0^{\ell+1} \cdot 4^\ell \cdot \frac{1}{\ell!} \cdot t_{j-\ell-1}^{-\ell/2}.$$

We shall show this by induction (with the base case, in which $\ell = 0$, being proven above).

$$\begin{aligned} &\Upsilon_{j-\ell-1}(C_{j-\ell-2}, t_{j-\ell-2}) \\ &\leq x^2 \zeta_0^{\ell+1} \cdot 4^{\ell-1} \cdot (\ell!)^{-1} \sum_{r_{j-\ell-2}} \sum_{C_{j-\ell-1}, D_{j-\ell-1}} \sum_{\Delta t_{j-\ell-1}} \frac{3}{\Delta t_j^3} \exp\left(\frac{-r_{j-\ell-2}^2}{4\Delta t_{j-\ell-1}}\right) \frac{1}{t_{j-1}^{1.5+\ell/2}} \\ &\leq x^2 \zeta_0^{\ell+1} \cdot 4^{\ell-1} (\ell!)^{-1} \sum_{\Delta t_{j-\ell-1}} \frac{1}{\Delta t_{j-\ell-1}^3 \cdot t_{j-\ell-1}^{1.5+\ell/2}} \left(2 \int_0^{\infty} 144 r_{j-\ell-2}^5 \exp\left(\frac{-r_{j-\ell-2}^2}{4\Delta t_{j-\ell-1}}\right) dr_{j-\ell-2}\right) \\ &= x^2 \zeta_0^{\ell+2} 4^\ell (\ell!)^{-1} \sum_{\Delta t_{j-\ell-1}} \frac{1}{t_{j-\ell-1}^{1.5+\ell/2}} \\ &\leq x^2 \zeta_0^{\ell+2} 4^{\ell+1} ((\ell+1)!)^{-1} \frac{1}{t_{j-\ell-2}^{(\ell+1)/2}}. \end{aligned}$$

The last inequality holds because

$$\sum_{\Delta t_{j-\ell-1}} \frac{1}{t_{j-\ell-1}^{1.5+\ell/2}} \leq 2 \int_{t_{j-\ell-2}}^{\infty} \frac{1}{t^{1.5+\ell/2}} dt \leq 4(\ell+1)^{-1} \frac{1}{t_{j-\ell-2}^{(\ell+1)/2}}.$$

This completes the induction. Finally, we have

$$\begin{aligned} \Upsilon_1 &= \sum_{C_1} \sum_{t_1} f_1 \frac{1}{t_1^{1.5}} \Upsilon_2(C_1, t_1) \\ &\leq x^2 \zeta_0^{j-1} 4^{j-2} ((j-2)!)^{-1} \sum_{C_1, t_1} \frac{1}{t_1^{j/2+3.5}} \cdot \exp \left(\frac{-3\|B-C_1\|_2^2 + \|A_1-C_1\|_2^2}{2t_1} \right). \end{aligned}$$

Next, let $\|B-C_1\|_1 = r_0$. By Equation 2.7, $\|B-C_1\|_2^2 \geq r_0^2/3$. By Equation 2.8, we have $\|B-C_1\|_2^2 + \|A_1-C_1\|_2^2 \geq x^2/6$. Therefore, we have $\|B-C_1\|_2^2 + \|A_1-C_1\|_2^2 \geq \frac{1}{6}(r_0^2 + x^2/2)$. We have

$$\begin{aligned} &\sum_{C_1, t_1} \frac{1}{t_1^{j/2+3.5}} \exp \left(\frac{-3(\|B-C_1\|^2 + \|A_1-C_1\|^2)}{2t_1} \right) \\ &\leq \sum_{t_1} \frac{1}{t_1^{j/2+3.5}} \sum_{\|B-C_1\|_1=r_0} \sum_{C_1} \exp \left(\frac{-3r_0^2}{12t_1} \right) \cdot \exp \left(\frac{-3x^2}{24t_1} \right) \\ &\leq \sum_{t_1} \frac{1}{t_1^{j/2+3.5}} \left(2 \int_0^\infty 4r_0^2 \exp \left(\frac{-r_0^2}{4t_1} \right) dr_0 \right) \cdot \exp \left(\frac{-3x^2}{24t_1} \right) \\ &\leq \sum_{t_1} \frac{1}{t_1^{j/2+3.5}} 16\sqrt{\pi} t_1^{1.5} \exp \left(\frac{-x^2}{8t_1} \right) \\ &\leq 30 \sum_{t_1} \frac{1}{t_1^{j/2+2}} \exp \left(\frac{-x^2}{8t_1} \right) \\ &\leq 60 \frac{\Gamma(j/2+1)}{(x^2/8)^{j/2+1}}. \end{aligned}$$

The third inequality holds because

$$\int_0^\infty r_0^2 \exp \left(-\frac{r_0^2}{4t_1} \right) dr_0 = 2\sqrt{\pi} t_1^{1.5}$$

The last inequality holds because

$$\int_0^\infty y^{-c} \exp \left(\frac{-x^2}{8y} \right) dy = \frac{8^{c-1}}{x^{2(c-1)}} \Gamma(c-1)$$

for any constant c and real number x .

We thus conclude that

$$\Upsilon_1 \leq \frac{x^2 \zeta_0^{j-1} 4^{j-2}}{(j-2)!} \cdot 60 \cdot \frac{\Gamma(\frac{j}{2} + 1)}{x^{2(j/2+1)}} 8^{j/2+1} \leq \frac{30(8\sqrt{2})^j \zeta_0^{j-1} \Gamma(\frac{j}{2} + 1)}{(j-2)! x^j}.$$

When the permutation is considered, we have

$$\begin{aligned} & \Pr[\exists t_1, \dots, t_j, C_1, \dots, C_j \Psi_{C_1, \dots, C_j}^{t_1, \dots, t_j} = 1] \\ & \leq j! \frac{30(8\sqrt{2})^j \zeta_0^{j-1} \Gamma(\frac{j}{2} + 1)}{(j-2)! x^j} \\ & \leq \frac{30j(j-1)(8\sqrt{2})^j \zeta_0^{j-1} \Gamma(\frac{j}{2} + 1)}{x^j} \\ & \leq \frac{(8\sqrt{2}\zeta_0)^j j^j}{x^j} \\ & \leq \left(\frac{8\sqrt{2}\zeta_0 j}{x} \right)^j \end{aligned}$$

By setting $\zeta = 8\sqrt{2}\zeta_0 < 210000$, our lemma follows. \square

Lemma 2.2.4 also helps us to analyze the scenario in which agents need to meet rather than to collide. This is summarized by the following corollary:

Corollary 2.2.5. *Let A_1, A_2, \dots, A_j , and B be points in \mathbf{Z}^3 such that $\|A_i - B\|_1 \geq x$ for all $i \leq j$. Let $S^1(A_1), \dots, S^j(A_j), S^{j+1}(B)$ be $j+1$ independent random walks that start at A_1, \dots, A_j, B respectively. Let $t = x^2$. Then the probability that all the walks S^1, \dots, S^j meet with S^{j+1} within time t is at most $\left(\frac{\zeta' j}{x}\right)^j$, where ζ' is a sufficiently large constant.*

Proof of Corollary 2.2.5. Notice first that if $S^{j+1}(B)$ and $S^i(A_i)$ meet at a time step t_0 , then there exists a point A'_i with $\|A'_i - A_i\|_1 = 1$ such that the walk $S^{i'}(A'_i)$ that mimics the moves of S^i at each step collides with S^{j+1} at time t_0 .

Therefore, a necessary condition for S^{j+1} to meet the rest of agents is that there exist $S^{1'}(A'_1), S^{2'}(A'_2), \dots, S^{j'}_{A'_j}$ such that

- $\|A'_i - A_i\|_1 = 1$ for all $i \leq j$.
- $S^{i'}$ mimics the moves of S^i at all steps for all $i \leq j$.
- S^{j+1} *collides* with all of $S^{1'}, \dots, S^{j'}$ before time t .

For any A'_1, \dots, A'_j , the collision probability is at most $\left(\frac{\zeta_j}{x}\right)^j$ by Lemma 2.2.4. The total number of possible j -tuples A'_1, \dots, A'_j is 7^j . By using a union bound, the probability there exists a j -tuple such that all j walks collide with S^{j+1} is at most $7^j \left(\frac{\zeta_j}{x}\right)^j$. The corollary follows. \square

We note that both Lemma 2.2.4 and Corollary 2.2.5 are useful only when x is large enough. This forms a barrier for analysis of close agents in our model. But as we will see, we can get around this issue by looking at a coupled diffusion process that possesses a different diffusion rule specifically designed for handling close agents.

Another important issue is the analysis on walks that are close to the boundary. For this, we show that the random walks will not behave significantly different (in terms of the desired bounds) when boundaries are added. We notice that similar results are presented in [92], but their results do not immediately translate to the building blocks we need here. The following is the major building block we need for our analysis:

Lemma 2.2.6. *Let A and B be two points in \mathcal{V}^3 such that $A - B = \vec{x}$ and the distance between A and any boundary is at least $40\|\vec{x}\|_1$. Consider two random walks $S^1(A)$*

and $S^2(B)$ that start at A and B respectively. Let \tilde{e}_t be the event that $S^1(A)$ and $S^2(B)$ will meet before time $t = \|\vec{x}\|_1^2$ and before either of them visits a boundary. Then $\Pr[\tilde{e}_{\|\vec{x}\|_1^2}] = \Omega(1/\|\vec{x}\|_1)$.

We next move to prove Lemma 2.2.6. Since we need to frequently compare bounded random walks with their unbounded counterparts, we use \mathbf{S} to represent unbounded walks and S to represent bounded walks in the rest of this section.

Our analysis consists of two steps. We first tackle a simpler problem, in which we need to understand the probability for a random walk starting from a point near the boundary to visit another point in \mathcal{V}^3 within a short time frame. We then utilize results from this scenario to prove Lemma 2.2.6.

Lemma 2.2.7. *Let $\mathcal{V}^3 = \{-n, \dots, n\}^3$. Let A and B be two points in \mathcal{V}^3 such that $A - B = \vec{x}$ and the distance under L_∞ norm between A and any boundary is at least $20\|\vec{x}\|_1$. Consider a random walk $S(A)$ that starts at A . Let e_t^1 be the event that $S(A)$ is at B at time t . Let e_t^2 be the event that $S(A)$ hits a boundary at or before t . When $t = \Theta(\|\vec{x}\|_1^2)$, we have $\Pr[e_t^1 \wedge \neg e_t^2] \geq c_0 p(t, \vec{x})$ for some constant c_0 .*

Proof. First, let us couple the random walk $S(A)$ with a standard unbounded random walk $\mathbf{S}(A)$ in the natural way. Let \hat{e}_t^1 be the event that $\mathbf{S}(A)$ is at B at time t and let \hat{e}_t^2 be the event that $\mathbf{S}(A)$ ever visits a boundary at or before time t . When $\neg e_t^2$ occurs, $\mathbf{S}(A)$ and $S(A)$ coincide and $\Pr[e_t^1 \wedge \neg e_t^2] = \Pr[\hat{e}_t^1 \wedge \neg \hat{e}_t^2]$.

On the other hand, we have

$$p(t, \vec{x}) = \Pr[\hat{e}_t^1 \wedge \hat{e}_t^2] + \Pr[\hat{e}_t^1 \wedge \neg \hat{e}_t^2].$$

Notice that in the event $\hat{e}_t^1 \wedge \hat{e}_t^2$, $\mathbf{S}(A)$ has to travel from the boundary to B within

a time interval shorter than t . The distance between the boundary and B is at least $19\|\vec{x}\|_1$. Together with the analytic form of $p(\cdot, \cdot)$ in Lemma 2.6.5, we have $\Pr[\hat{e}_t^1 \wedge \hat{e}_t^2 = 1] \leq \max_{\|\vec{y}\|_1 \geq 19\|\vec{x}\|_1} p(t, \vec{y})$. Therefore, $\Pr[\hat{e}_t^1 \wedge \neg \hat{e}_t^2] \geq p(t, \vec{x}) - \max_{\|\vec{y}\|_1 \geq 19\|\vec{x}\|_1} p(t, \vec{y})$. Finally, we have

$$\Pr[e_t^1 \wedge \neg e_t^2] = \Pr[\hat{e}_t^1 \wedge \neg \hat{e}_t^2] \geq p(t, \vec{x}) - \max_{\|\vec{y}\|_1 \geq 19\|\vec{x}\|_1} p(t, \vec{y}) \geq \frac{1}{2}p(t, \vec{x}).$$

We may use the analytic form of the function $p(\cdot, \cdot)$ (Lemma 2.6.5) for $t = \Theta(\|\vec{x}\|_1^2)$ to verify the last inequality. \square

Proof of Lemma 2.2.6. Let X be the number of collisions between S^1 and S^2 that are before time t and before either of them visits a boundary. Also let $\hat{e}(S_t)$ be the event that the random walk S ever visits a boundary at or before time t . We have

$$\begin{aligned} & \mathbb{E}[X] \\ &= \sum_{t \leq \|\vec{x}\|_1^2} \Pr[(S_t^1 = S_t^2) \wedge (\neg \hat{e}(S_t^1) \wedge \neg \hat{e}(S_t^2))] \\ &= \sum_{t \leq \|\vec{x}\|_1^2} \sum_{C \in \mathcal{V}} \Pr[(S_t^1 = C) \wedge \neg \hat{e}(S_t^1)] \Pr[(S_t^2 = C) \wedge \neg \hat{e}(S_t^2)] \quad (\text{two walks are independent}) \\ &\geq \sum_{t \leq \|\vec{x}\|_1^2} \sum_{C: \|C-A\|_1 \leq \|\vec{x}\|_1} \Pr[(S_t^1 = C) \wedge \neg \hat{e}(S_t^1)] \Pr[(S_t^2 = C) \wedge \neg \hat{e}(S_t^2)] \end{aligned}$$

The last inequality holds because we only focus on a subset of \mathcal{V}^3 . Since $\|C - A\|_1 \leq \|\vec{x}\|_1$ and $\|A - B\|_1 \leq \|\vec{x}\|_1$, we have $\|C - B\|_1 \leq 2\|\vec{x}\|_1$.

By Lemma 2.2.7,

$$\Pr[(S_t^1 = C) \wedge \neg \hat{e}(S_t^1)] \geq \frac{1}{2}p(t, A - C) \quad \text{and} \quad \Pr[(S_t^2 = C) \wedge \neg \hat{e}(S_t^2)] \geq \frac{1}{2}p(t, B - C)$$

We now have

$$\begin{aligned}
& \sum_{t \leq \|\vec{x}\|_1^2} \sum_{C: \|C-A\|_1 \leq \|\vec{x}\|_1} \Pr[(S_t^1 = C) \wedge \neg \hat{e}(S_t^1)] \Pr[(S_t^2 = C) \wedge \neg \hat{e}(S_t^2)] \\
& \geq \sum_{1 \leq t \leq \|\vec{x}\|_1^2} \sum_{C: \|C-A\|_1 \leq \|\vec{x}\|_1} \frac{1}{2} p(t, A-C) \frac{1}{2} p(t, B-C) \quad (\text{by Lemma 2.2.7}) \\
& = \Omega \left(\sum_{1 \leq t \leq \|\vec{x}\|_1^2} \|\vec{x}\|_1^3 \min_{C: \|C-A\|_1 \leq \|\vec{x}\|_1} \{p(t, C-A)p(t, B-C)\} \right) \\
& = \Omega(1/\|\vec{x}\|_1).
\end{aligned}$$

The last equality can be shown by using the analytic form of $p(\cdot, \cdot)$ again (Lemma 2.6.5) and the fact that $\|A-C\|_2$ and $\|B-C\|_2$ are in $O(\|\vec{x}\|_2)$.

Next, let us compute $E[X | X \geq 1]$, i.e., the expected number of collisions when they collide at least once. Upon the first time S^1 and S^2 collide (before either of them visit the boundary), we couple S^1 and S^2 with two unbounded random walks \mathbf{S}^1 and \mathbf{S}^2 in the natural way respectively. The expected number of collisions between \mathbf{S}^1 and \mathbf{S}^2 for t steps (when they start at the same point) is an upper bound on $E[X | X \geq 1]$. On the other hand, we may couple \mathbf{S}^1 and \mathbf{S}^2 with a single random walk \mathbf{S} in the way described in Lemma 2.2.7 so that the expected number of collisions between \mathbf{S}^1 and \mathbf{S}^2 is the expected number of times \mathbf{S} returns to the point where it starts at.

Finally, the expected number of return for an unbounded random walk is a constant (which can be derived from $\sum_{t \geq 0} p(t, \vec{0})$, where $p(\cdot, \cdot)$'s analytic form is in Theorem 2.6.5). Therefore, $E[X | X \geq 1] = O(1)$. Now since $E[X] = E[X | X \geq 1] \Pr[X \geq 1]$. Therefore, $\Pr[\tilde{e}_{\|\vec{x}\|_1^2}] = \Pr[X \geq 1] = \Omega(1/\|\vec{x}\|_1)$.

□

2.3 Lower bound

Let us first state our lower bound result more precisely as follows.

Theorem 2.3.1. *Let a_1, \dots, a_m be placed uniformly at random on \mathcal{V}^3 such that $1600n \log^2 n \leq m \leq n^3$. Let $\ell_2 = \sqrt{n^3/m}$. For sufficiently large n , the diffusion time T satisfies the following inequality*

$$\Pr[T \leq \frac{1}{81} \ell_2 n \log^{-29} n] \leq \exp(-\log n \log \log n).$$

We use a *local analysis* to prove our lower bound. The key idea is that under uniform distribution of agents, the extent any particular infected agent can spread the virus within a small time increment is confined to a small neighborhood with high probability. By gluing together these local estimates, we can approximate the total diffusion time.

To explain our local analysis, assume we start with an arbitrary infected agent, say a_1 . Let us also assume, for simplicity, that all the other uniformly distributed agents are uninfected. Consider the scenario within a small time increment, say Δt . During this time increment the agent a_1 infects whoever it meets in the small neighborhood that contains its extent of movement. The newly infected agents then continue to move and infect others. The size of the final region that contains all the infected agents at Δt then depends on the rate of transmission and the extent of movement of all of the infected agents. In particular, if Δt is small enough, the expected number of transmissions performed by a_1 is less than one; even if it infects another agent, the number of infections it causes within the *same* Δt is also less than one, and so on. The net effect is an eventual dying-out of this “branching process”

(which we later model by what we call a diffusion tree), which localizes the positions of all infected agents at time Δt to a small neighborhood around the initial position of a_1 .

As it may not be clear as we go through our proofs, we briefly review the main methodologies in obtaining lower bound results in related work, and point out their relation to our analysis and difficulties in directly applying them to higher dimensions. (Some readers may wish to skip these next paragraphs all together; for others, who would like a more thorough discussion that unavoidably requires more technical details, we devote Chapter 2.7 to more details.) Two potential existing methods arise in [5, 71] and [92]. The former analyzes the growth rate of the size of the *total* infected region; an upper bound on this growth rate translates to a lower bound for the diffusion time. The latter work, focusing on $d = 1, 2$, uses an “island diffusion rule”, which essentially speeds up infection by allowing infections to occur immediately on connected components in an underlying graph where edges are based on the distance between agents. This approach avoids handling the issue of the meeting time of random walks when they are very close, a regime where asymptotic results such as Lemma 2.2.3 and 2.2.4 may not apply, while still providing a way to bound the diffusion time by arguing about the low probability of interaction among different “islands”.

The results in [5, 71] are not directly applicable in our setting because the growth rate they obtain is linear in time, as a result of their assumption of constant agent density in an infinite space, in contrast to our use of a size parameter n that scales with the agent density. It is fairly simple to see that blindly applying a linear growth

rate to our setting of $o(1)$ density is too crude. On the other hand, analyzing how agent density affects the growth rate is a potentially feasible approach but certainly not straightforward.

Our approach more closely follows [92]. The main limitation of [92], when applied to higher dimension, is how to control the interaction among islands. If islands interact too often, because they are too close together, the argument, which is based on a low probability of interaction, breaks down. However, if one parametrizes islands to prevent such interaction, then the bound that can be obtained are too weak. In Chapter 2.7.1 we provide further details arguing that for $d > 2$ this constraint ultimately limits the analysis for the case of $o(1)$ density. We attempt to remedy the problem by using islands as an intermediate step to obtain local estimates of the influence of each initially infected agent over small periods of time. This analysis involves looking at a branching process representing the spread of the infection, significantly extending the approach of [92].

2.3.1 Local diffusion problem

This subsection focuses on the local analysis as discussed above. In Chapter 2.3.2, we will proceed to discuss how to utilize this analysis to get the lower bound in Theorem 2.3.1. As discussed in the last section, the two main difficulties in our analysis are: 1) our probabilistic estimates for the meeting time/position of multiple random walks are only useful asymptotically; 2) walks near the boundary introduce further analytical complication. To begin with, the following definition serves to handle the second issue:

Definition 2.3.2 (Interior region). *The interior region $\mathfrak{V}(r)$ parameterized by r is the set of lattice points in \mathcal{V}^3 that have at least L_∞ -distance r to the boundary.*

For any point $P \in \mathcal{V}^3$, define $\mathbb{B}(P, x) = \{Q \in \mathcal{V}^3 : \|Q - P\|_\infty \leq x\}$ as the x -ball of neighborhood of P under L_∞ -norm. The following proposition is our major result in this subsection.

Proposition 2.3.3. *Consider a diffusion following Definition 2.1.1. Let S_0 be the initial position of the only infected agent a_1 at time 0, and \mathcal{W} be an arbitrary subset of lattice points in $\mathfrak{V}(20\ell_2 \log n)$, where $\ell_2 = \sqrt{n^3/m}$. Denote $\Delta t = \ell_2^2 \log^{-28} n$. Define the binary random variable $b(\mathcal{W})$ as follows:*

- If $S_0 \in \mathcal{W}$: $b(\mathcal{W})$ is set as 1 if and only if all the infected agents at time Δt can be covered by the ball $\mathbb{B}(S_0, 9\ell_2 \log n)$.
- If $S_0 \notin \mathcal{W}$: $b(\mathcal{W}) = 1$.

We have

$$\Pr[b(\mathcal{W}) = 1] \geq 1 - \exp(-5 \log n \log \log n) \quad (2.12)$$

The proposition yields that with high probability, all the infected agents lie within a neighborhood of distance $\tilde{O}(\ell_2)$ at time $\tilde{O}(\ell_2^2)$. The variable ℓ_2 is chosen such that the expected number of infections spanned by an initially infected agent a_1 within $\tilde{O}(\ell_2^2)$ units of time and a neighborhood of $\tilde{O}(\ell_2)$ distance is $O(1)$. This can be seen by solving $m(\ell_2/n)^3 \times (1/\ell_2) = \tilde{O}(1)$, where $m(\ell_2/n)^3$ is the expected number of agents in a cube of size $\ell_2 \times \ell_2 \times \ell_2$, and $\tilde{O}(1/\ell_2)$ is the meeting probability within time $\tilde{O}(\ell_2^2)$ between any pair of random walks with initial distance ℓ_2 (see Lemma 2.2.3). This choice of ℓ_2 appears to be the right threshold for our analysis. Indeed, a larger scale

than ℓ_2 would induce a large number of infections made by a_1 , and also subsequent infections made by newly infected agents, with an exploding affected region as an end result. On the other hand, a smaller scale than ℓ_2 would degrade our lower bound. This is because the diffusion time is approximately of order n/ℓ_2 , the number of spatial steps to cover \mathcal{V}^3 , times ℓ_2^2 , the time taken for each step, equaling $n\ell_2$. Hence a decrease in ℓ_2 weakens the bound¹.

Secondly, we introduce \mathcal{W} in Proposition 2.3.3 to avoid the case when S_0 is close to the boundary. As we have mentioned, such boundary conditions often complicate random walk analysis. Although the impact of the boundary's presence has been addressed (e.g., [26, 92]), existing results are not fully satisfactory. For example, when two simple random walks S^1 and S^2 start near the boundary, only a lower bound for the probability that two walks meet within a specific number of time steps is available ([92]); we do not know of an upper bound counterpart. We arrange our proof so that it is sufficient to analyze the diffusion pattern of a virus when it starts far from the boundary. Finally, we note that no effort has been made to optimize the exponent 28 in Δt 's definition.

We briefly explain how our global lower bound can be readily obtained from Proposition 2.3.3, which is a strong characterization of the local growth rate of infection region size. Imagine the following evolution. Starting with a single infected agent, with high probability the infection spreads to a ball of radius at most $9\ell_2 \log n$ in Δt time units. At this time point, the newly infected agents *inside* the ball con-

¹In the case of general d -dimensional space, ℓ_2 is chosen such that $m(\ell_2/n)^d \times (1/\ell_2^{d-2}) = \tilde{O}(1)$, giving $\ell_2 = \sqrt{n^d/m}$. Throughout the chapter such d -dimensional analog can be carried out in similar fashion, but for ease of exposition we shall not bring up these generalizations and will focus on the 3-dimensional case.

tinue to spread the virus to neighborhoods of size at most $9\ell_2 \log n$, again with high probability. This gives an enlarged area of infection with radius at most $18\ell_2 \log n$. Continuing in this way, the lower bound in Theorem 2.1.3 is then the time for the infection to spread over \mathcal{V}^3 . This observation will be made rigorous in the next subsection.

The rest of this subsection is devoted to the proof of Proposition 2.3.3. It consists of two main steps. First, we need to estimate the expected number of infections done by a single initially infected agent within distance $9\ell_2 \log n$ and time increment Δt . Second, we iterate to consider each newly infected agent. The analysis requires the condition that the global configuration behaves “normally”, a scenario that occurs with suitably high probability, as we show. We call this condition “good behavior”, which is introduced through the several definitions below:

Definition 2.3.4. (Island, [92]) *Let $A = \{a_1, \dots, a_m\}$ be the set of agents in \mathcal{V}^3 . For any positive integer $\gamma > 0$, let $G_t(\gamma)$ be the graph with vertex set A such that there is an edge between two vertices if and only if the corresponding agents are within distance γ (under L_1 -norm) at time t . The island with parameter γ of an agent $a_i \in A$ at time step t , denoted by $\text{Isd}_t(a_i, \gamma)$ is the connected component of $G_t(\gamma)$ containing a_i .*

Definition 2.3.5 (Good behavior). *Let $\ell_1 = nm^{-1/3}$. For $1 \leq i \leq (\ell_2/\ell_1) \log^{-3} n$, define $\mathcal{B}_i(P) = \mathbb{B}(P, i\ell_1 \log^{-1} n)$ and let $\partial\mathcal{B}_i(P) = \mathcal{B}_i(P) - \mathcal{B}_{i-1}(P)$. For any $P \in \mathcal{V}^3$, define $m_i(P) = \frac{(\log^5 n)|\partial\mathcal{B}_i(P)|m}{(2n+1)^3}$. Let us define the following binary random variables:*

- **Good density.** *Let $\{D_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $D_t = 1$ if and only if for all $P \in \mathcal{V}^3$ and all $i \leq (\ell_2/\ell_1) \log^{-3} n$, the number*

of agents in $\partial\mathcal{B}_i(P)$ is at most $m_i(P)$, for all time steps up to t . We say the diffusion process has the good density property at time t if $D_t = 1$.

- **Small islands.** Let $\{E_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $E_t = 1$ if and only if $|\text{Isd}_s(a_j, \ell_1 \log^{-1} n)| \leq 3 \log n$ for all $a_j \in A$ and $0 \leq s \leq t$. We say that the diffusion process has the small islands property at time t if $E_t = 1$.
- **Short travel distance.** Let $\{L_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $L_t = 1$ if and only if for all $i \in [m]$ and all $t_1 < t_2 \leq t$ with $t_2 - t_1 \leq \ell_2^2 \log^{-12} n$, we have $\|S_{t_1}^i - S_{t_2}^i\|_1 \leq 3\ell_2 \log^{-4} n$. We say the process has the short travel distance property at time t if $L_t = 1$.

Finally, let $G_t = D_t \times E_t \times L_t$, and say the diffusion process behaves well at time t if $G_t = 1$. We also focus on $t \leq n^{2.5}$ and define the random variable $G = G_{n^{2.5}}$.

The value $n^{2.5}$ in the definition is chosen such that it lies well beyond our lower bound for the case $m < n^3$, but is small enough for our forthcoming union bound. By using properties of random walks and techniques derived in [92], we have

Lemma 2.3.6. *Let $A = \{a_1, \dots, a_m\}$ be agents that are distributed uniformly in \mathcal{V}^3 at $t = 0$. For sufficiently large n , we have $\Pr[G = 1] \geq 1 - \exp(-6 \log n \log \log n)$.*

Proof of Lemma 2.3.6. We first show the good density property holds with high probability. For any specific time $t \leq n^{2.5}$, all the agents are uniformly distributed due to stationarity. For an arbitrary $P \in \mathcal{V}^3$, and $i \leq (\ell_2/\ell_1) \log^{-3} n$, define $Y(t, P, i)$ as the number of agents that are in $\partial\mathcal{B}_i(P)$ at time t . Notice that $\mathbb{E}[Y(t, P, i)] =$

$|\partial\mathcal{B}_i(P)|m/(2n+1)^3$ and $m_i(P) \geq 6\mathbb{E}[Y(t, P, i)]$. By Chernoff bounds (e.g., the second part of Theorem A.1.1),

$$\begin{aligned}
\Pr[Y(t, P, i) \geq m_i] &\leq 2^{-m_i} \\
&\leq \exp(-0.65m_i) \\
&\leq \exp\left(0.65\frac{\ell_1^3 \log^{-3} n}{n^3} m \log^5 n\right) \\
&\leq \exp(-0.65 \log^2 n)
\end{aligned}$$

for sufficiently large n . Next, by a union bound,

$$\begin{aligned}
\Pr[D_t = 0] &\leq \sum_{t, P, i} \Pr[Y(t, P, i) \geq m_i] \\
&\leq (n^{2.5}(2n+1)^3(\log^{-3} n)\ell_2/\ell_1) \exp(-0.65 \log^2 n) \\
&\leq \exp(-\frac{1}{2} \log^2 n).
\end{aligned}$$

Therefore, we have $\Pr[D_t = 1] \geq 1 - \exp(-\frac{1}{2} \log^2 n)$.

To show the diffusion process has the small islands property with high probability, we mimic the proof of Lemma 6 in [92]. Let B_k be the event that there exists an island with parameter $\gamma = \ell_1 \log^{-1} n$ that has at least k agents. The quantity $\Pr[B_k]$ is upper bounded by the probability that $G_t(\gamma)$ contains a tree of k vertices of A as a subgraph. Since k^{k-2} is the number of unrooted labeled trees on k nodes, and γ^3/n^3 is an upper bound to the probability that a given agent lie within distance γ from another given agent, we have that

$$\Pr[B_k] \leq \binom{m}{k} k^{k-2} \left(\frac{\gamma^3}{n^3}\right)^{k-1} \leq \left(\frac{em}{k}\right)^k \cdot k^{k-2} \left(\frac{1}{m \log^3 n}\right)^{k-1} = \frac{e^k m}{k^2} \cdot (\log n)^{-3(k-1)}.$$

By setting $k = 3 \log n + 1$, we have $\Pr[B_k] \leq \exp\{-7 \log n \cdot \log \log n\}$. Finally, we apply a union bound across all agents and all time steps. Hence $\Pr[E_t = 1] > 1 - n^{2.5} m \exp(-7 \log n \log \log n)$.

Finally, consider the short travel distance property. For any fixed $i \in [m]$ and $t_1 \leq t_2 \leq n^{2.5}$ such that $t_2 - t_1 \leq \ell_2^2 \log^{-12} n$, we have $\Pr[\|S_{t_1}^i - S_{t_2}^i\|_\infty \geq \ell_2 \log^{-4} n] \leq \exp(-\log^2 n)$ by Lemma 2.6.2. There is a factor of 3 lost when we translate the metric from L_∞ -norm to L_1 -norm. The total number of possible i , t_1 , and t_2 are mn^5 . Next we may apply a union bound across all these possible i , t_1 , and t_2 triples. We have $\Pr[L_t = 0] \leq mn^5 \exp(-\log^2 n)$. The lemma follows by combining the three results together with one more union bound. \square

With this global “good behavior”, we have the following estimate:

Lemma 2.3.7. *Let $A = \{a_1, \dots, a_m\}$ be agents that are distributed in \mathcal{V}^3 in such a way that $D_0 = 1$. Let S^1, S^2, \dots, S^m be their corresponding random walks. Consider an arbitrary agent a_j with $S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)$. Let $\{a_{i_1}, \dots, a_{i_k}\}$ be the set of agents outside $\mathcal{B}_1(S_0^j)$ at time 0. Define $X_{j,\ell}$ as the indicator random variable that represents whether the agents a_j and a_{i_ℓ} meet within time $[0, \Delta t]$. We have*

$$\mathbb{E} \left[\sum_{\ell \leq k} X_{j,\ell} \middle| D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] < \log^{-3} n.$$

Proof. First, notice that the number of lattice points in $\partial\mathcal{B}_i(P)$ satisfies

$$|\partial\mathcal{B}_i| = |\mathcal{B}_i| - |\mathcal{B}_{i-1}| \leq (2i\ell_1 \log^{-1} n)^3 - (2(i-1)\ell_1 \log^{-1} n)^3 \leq 24i^2\ell_1^3 \log^{-3} n.$$

We may also similarly show that

$$|\partial\mathcal{B}_i| \geq i^2\ell_1^3 \log^{-3} n.$$

Let $q = (\ell_2/\ell_1) \log^{-3} n$. For each $i \in [q]$, write $\mathcal{B}_i = \mathcal{B}(S_0^j)$, $\partial\mathcal{B}_i = \partial\mathcal{B}_i(S_0^j)$, and $m_i = m_i(S_0^j)$. We want to estimate the meeting probability and hence the expected number of infections for each $i \in [q]$.

First, let us consider the agents outside the ball \mathcal{B}_q . The probability that any specific agent initially outside \mathcal{B}_q ever travels into the ball $\mathbb{B}(S_0^j, \ell_2 \log^{-4} n)$ within time $\ell_2^2 \log^{-12} n$ is at most $\exp(-\Omega(\log^3 n))$ (by, e.g., Lemma 2.6.2 in Chapter 2.6). On the other hand, the probability that S^j ever travels out of $\mathbb{B}(S_0^j, \ell_2 \log^{-4} n)$ is also $\exp(-\Omega(\log^3 n))$. For these two agents to meet, at least one of these events has to occur. Therefore, with probability $\exp(-\Omega(\log^3 n))$ S^j will meet an agent initially outside \mathcal{B}_q . This leads to

$$\mathbb{E} \left[\sum_{\substack{i_{k'}: S_0^{i_{k'}} \notin \mathcal{B}_q \\ \text{the set of agents initially} \\ \text{outside } \mathcal{B}_q}} X_{j,k'} \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \leq m \exp(-\Omega(\log^3 n)).$$

Let us next focus on agents inside \mathcal{B}_q . Fix an arbitrary $a_{i_\ell} \in \partial \mathcal{B}_i$. Let e^j and e^ℓ represents the events that S^j and S^{i_ℓ} ever visit a boundary before time $\ell_2^2 \log^{-12} n$ respectively. Again by Lemma 2.6.2, $\Pr[e^j \vee e^\ell \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] = \exp(-\Omega(\log^3 n))$. We now have

$$\begin{aligned} & \mathbb{E}[X_{j,\ell} \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \\ &= \Pr[X_{j,\ell} = 1; \neg e^j \wedge \neg e^\ell \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \\ & \quad + \Pr[X_{j,\ell} = 1; e^j \vee e^\ell \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \\ &\leq \Pr[X_{j,\ell} = 1; \neg e^j \wedge \neg e^\ell \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] + \exp(-\Omega(\log^3 n)). \end{aligned}$$

To compute

$$\begin{aligned} & \Pr[X_{j,\ell}; \neg e^j \wedge \neg e^\ell \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \\ &= \Pr \left[\left\{ \exists t_0 \leq \frac{\ell_2^2}{\log^{12} n} : \|S_{t_0}^j - S_{t_0}^{i_\ell}\|_1 \leq 1 \right\} \wedge (\neg e^j \wedge \neg e^\ell) \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right], \end{aligned}$$

we couple S^j and S^{i_ℓ} with unbounded walks \mathbf{S}^j and \mathbf{S}^{i_ℓ} starting at the same positions at $t = 0$ in the natural way. Before the pair of bounded walks visit the boundary,

they coincide with their unbounded counterparts. Therefore, we have

$$\begin{aligned}
& \Pr \left[\left\{ \exists t_0 \leq \frac{\ell_2^2}{\log^{12} n} : \|S_{t_0}^j - S_{t_0}^{i_\ell}\|_1 \leq 1 \right\} \bigwedge (\neg e^j \wedge \neg e^\ell) \middle| D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \\
& \leq \Pr \left[\exists t_0 \leq \frac{\ell_2^2}{\log^{12} n} : \|\mathbf{S}_{t_0}^j - \mathbf{S}_{t_0}^{i_\ell}\|_1 \leq 1 \middle| D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \\
& = O\left(\frac{1}{(i-1)\ell_1}\right) \quad (\text{Corollary 2.2.5})
\end{aligned}$$

We thus have $\mathbb{E}[X_{j,\ell} | D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \leq \frac{C_0}{(i-1)\ell_1}$ for some constant C_0 .

Next, we estimate $E_i \equiv \mathbb{E}[\sum_{\ell: S_0^{i_\ell} \in \partial \mathcal{B}_i} X_{j,\ell} \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)]$ as

$$E_i \leq \frac{C_0 m_i}{(i-1)\ell_1} = \frac{C_0}{(i-1)\ell_1} \cdot \frac{|\partial \mathcal{B}_i| m \log^5 n}{8n^3} = \frac{C_0 (3i^2 \ell_1^3 \log^{-3} n) m \log^5 n}{(i-1)\ell_1 n^3} \leq \frac{6C_0 i m \ell_1^2 \log^2 n}{n^3}.$$

The first inequality holds because $D_0 = 1$ and m_i is an upper bound for the number of agents in $\partial \mathcal{B}_i$ (for all i).

$$\begin{aligned}
& \mathbb{E}[\sum_{\ell \leq k} X_{j,\ell} \mid D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)] \\
& \leq \sum_{i \leq q} E_i + \underbrace{m \times \exp(-\Omega(\log^3 n))}_{\text{upper bound for those outside } B_q} \\
& \leq \left(\sum_{2 \leq i \leq q} i \right) \frac{6C_0 m \ell_1^2 \log^2 n}{n^3} + \exp(-\Omega(\log^2 n)) \\
& < \frac{6C_0 q^2 m \ell_1^2 \log^2 n}{n^3} + \exp(-\Omega(\log^2 n)) \\
& = 6C_0 \log^{-4} n + \exp(-\Omega(\log^2 n)) \\
& < \log^{-3} n
\end{aligned}$$

for sufficiently large n as $m < n^3$. □

Lemma 2.3.7 says that if the initial distribution of agents possesses good behavior, then one can ensure that the expected number of direct infections on far-away agents

is small. For agents close to the initially infected agents, we instead utilize the concept of islands, which is also deeply related to the subsequent virus spreading behavior. Now we formally introduce a new diffusion process with a modified “island diffusion” rule. It is easy to see that this new diffusion process can be naturally coupled with the original diffusion process (evolving with Definition 2.1.1) by using the same random walks in the same probability space.

Definition 2.3.8 (Diffusion process with island diffusion rule). *Consider a diffusion process in which m agents are performing random walks on \mathcal{V}^3 . An uninfected agent a_j becomes infected at time t if one of the following conditions holds:*

1. *it meets a previously infected agent at time t . For convenience, we say a_j is directly infected if it is infected in this way.*
2. *it is inside $\text{Isd}_t(a_i, \ell_1 \log^{-1} n)$ where a_i is directly infected at time t .*

This coupled process is different from the diffusion models introduced in [91, 92, 101]. In our formulation, an island is infected only if meeting occurs between one uninfected and one previously infected agent. In [91, 92, 101] (using our notations), an island is infected once it contains a previously infected agent. As a result, infections occur less frequently in our model than the models in [91, 92, 101]. This difference is the key to getting a tight lower bound for dimensions higher than 2. More precisely, our infection rule allows us to build a terminating branching process, or what we call “diffusion tree” in the following definition, whose generations are defined via the infection paths from the source. The termination of this branching process constrains the region of infection to a small neighborhood around the source with a probability

of larger order than obtained in [92]. This in turn leads to a tighter global lower bound.

Definition 2.3.9 (Diffusion tree). *Let $\mathcal{W} \subseteq \mathfrak{V}(2\ell_2 \log n)$ be a subset of lattice points. Consider a diffusion, following the island diffusion rule, that starts with an initially infected island $\text{Isd}_0(\mathbf{a}_1, \ell_1 \log^{-1} n)$. Recall that S_0^1 denotes \mathbf{a}_1 's position at $t = 0$. The diffusion tree Tr with respect to \mathcal{W} has the following components:*

1. *If $S_0^1 \notin \mathcal{W}$, $\text{Tr} = \emptyset$.*
2. *If $S_0^1 \in \mathcal{W}$,*
 - *The root of Tr is a dummy node r .*
 - *The children of r are all the agents in $\text{Isd}_0(\mathbf{a}_1, \ell_1 \log^{-1} n)$.*
 - *$\mathbf{a}_{\ell'}$ is a child of \mathbf{a}_ℓ ($\mathbf{a}_{\ell'} \in \text{child}(\mathbf{a}_\ell)$) if $\mathbf{a}_{\ell'}$ is infected by \mathbf{a}_ℓ before time Δt .*
 - *$\mathbf{a}_{\ell'}$ is a direct child of \mathbf{a}_ℓ ($\mathbf{a}_{\ell'} \in \text{dchild}(\mathbf{a}_\ell)$) if $\mathbf{a}_{\ell'} \in \text{child}(\mathbf{a}_\ell)$ and it is directly infected by \mathbf{a}_ℓ .*

For technical reasons, if $\mathbf{a}_{\ell'}$ is not in Tr , we let $\text{child}(\mathbf{a}_\ell) = \emptyset$ and $\text{dchild}(\mathbf{a}_\ell) = \emptyset$.

Notice that the diffusion tree Tr stops growing after Δt steps.

We refer the root of the tree as the 0th level of the tree and count levels in the standard way. The height of the tree is the number of levels in the tree. Note that diffusion tree defined in this way can readily be interpreted as a branching process (See, e.g., Chapter 0 in [109]), in which the j th generation of the process corresponds with the j th level nodes in Tr .

Next we incorporate the good behavior variable G_t with diffusion tree. The motivation is that, roughly speaking, consistently good behavior guarantees a small number of infections, or creation of children, at each level. This can be seen through Lemma 2.3.7.

Definition 2.3.10 (Stopped diffusion tree). *Consider a diffusion process with island diffusion rule, and let $T(\ell)$ be the time that a_ℓ becomes infected in the process. The stopped diffusion tree Tr' (with respect to a_i and \mathcal{W}) is a subtree of Tr induced by the set of vertices $\{a_\ell : a_\ell \in \text{Tr} \wedge G_{T(\ell)} = 1\}$. We write $a_\ell \in \text{child}'(a_{\ell'})$ if $a_\ell \in \text{child}(a_{\ell'})$ and $a_\ell \in \text{Tr}'$. Similarly, $a_\ell \in \text{dchild}'(a_{\ell'})$ if $a_\ell \in \text{dchild}(a_{\ell'})$ and $a_\ell \in \text{Tr}'$.*

Note that the definition of stopped diffusion tree involves global behavior of the whole diffusion process due to the introduction of G_t . On the other hand, $\text{Tr} = \text{Tr}'$ with overwhelming probability, so we can translate the properties of Tr' back to Tr easily.

We next show two properties of the (stopped) diffusion trees, one on the physical propagation of children relative to their parents and one on the tree height. These are our main ingredients for proving Proposition 2.3.3. The properties are in brief:

1. If a_ℓ is a child of $a_{\ell'}$ in the stopped diffusion tree Tr' , $\|S_{T(\ell)}^\ell - S_{T(\ell')}^{\ell'}\|_\infty$ is $\tilde{O}(\ell_2)$.
2. The height of the stopped diffusion tree Tr' is $\tilde{O}(1)$ with high probability.

Proving the first item requires the following notion:

Definition 2.3.11 (Generation distance). *Consider the diffusion tree Tr with respect to \mathcal{W} . Let a_ℓ be an arbitrary agent and let $a_{\ell'}$ be its parent on Tr . The generation*

distance of a_ℓ with respect to Tr is

$$\mathfrak{d}_\ell = \begin{cases} \|S_{T(\ell)}^\ell - S_{T(\ell')}^{\ell'}\|_1 & \text{if } a_\ell \text{ is in } \text{Tr} \text{ and is at the 2nd or deeper level} \\ 0 & \text{otherwise.} \end{cases} \quad (2.13)$$

In other words, the generation distance between a_ℓ and $a_{\ell'}$ is the distance between where a_ℓ and $a_{\ell'}$ were infected. The generation distance of a_ℓ with respect to Tr' is \mathfrak{d}'_ℓ , which is set to be \mathfrak{d}_ℓ if a_ℓ is in Tr' and 0 otherwise.

With this notion, we can derive the following lemma:

Lemma 2.3.12. *Consider the stopped diffusion tree with respect to \mathcal{W} that starts with an infected island $\text{Isd}_0(a_i, \ell_1 \log^{-1} n)$. For an agent a_ℓ in Tr' , $\mathfrak{d}'_\ell \leq 4\ell_2$.*

Proof. We focus on the non-trivial case that $S_0^i \in \mathcal{W}$ and that a_ℓ is at the 2nd level of Tr' or deeper. Suppose that the diffusion process behaves well up to time $T(\ell)$ i.e. $G_{T(\ell)} = 1$. Let $a_{\ell'}$ be the parent of a_ℓ on Tr' . By the construction of Tr' , there exists an $a_{\ell''} \in \text{dchild}'(a_{\ell'})$ (possibly a_ℓ itself) such that

- $a_\ell \in \text{Isd}_{T(\ell'')}(a_{\ell''}, \ell_1 \log^{-1} n)$.
- $T(\ell) = T(\ell'')$ i.e., a_ℓ and $a_{\ell''}$ get infected at the same time due to the island diffusion rule.
- $\|S_{T(\ell)}^{\ell'} - S_{T(\ell'')}^{\ell''}\|_1 \leq 1$ i.e., $a_{\ell''}$ gets infected because it meets an infected agent.

By the triangle inequality,

$$\mathfrak{d}'_\ell = \|S_{T(\ell')}^{\ell'} - S_{T(\ell)}^\ell\|_1 \leq \|S_{T(\ell')}^{\ell'} - S_{T(\ell'')}^{\ell''}\|_1 + \|S_{T(\ell'')}^{\ell''} - S_{T(\ell)}^\ell\|_1.$$

Note that $\|S_{T(\ell')}^{\ell''} - S_{T(\ell'')}^{\ell''}\|_1 \leq 3\ell_2 \log^{-4} n + 1 \leq \ell_2$ (short travel distance property) and $\|S_{T(\ell'')}^{\ell''} - S_{T(\ell)}^{\ell}\|_1 \leq (\ell_1 \log^{-1} n)(3 \log n) = 3\ell_1 \leq 3\ell_2$ (small island property) since $G_{T(\ell)} = 1$. Finally, the case when $G_{T(\ell)} = 0$ is trivial, and the lemma follows. \square

Next we show that with high probability the height of the stopped diffusion tree is $\tilde{O}(1)$. Using standard notation, we let $\{\mathcal{F}_t\}_{t \geq 0}$ be the σ -algebra, or filtration, generated up to time t , i.e., \mathcal{F}_t encodes all the information regarding the diffusion process up to t . The special instance \mathcal{F}_0 is used to describe the initial positions of the agents.

The main property of stopped diffusion tree that we need is the following:

Lemma 2.3.13. *Consider a diffusion process with the island diffusion rule. Let a_ℓ be an arbitrary agent with infection time $T(\ell)$. We have*

$$\mathbb{E} \left[|\text{dchild}'(a_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \leq \log^{-3} n, \quad (2.14)$$

where $\text{dchild}'(\cdot)$ is defined for a stopped diffusion tree with respect to an arbitrary set $\mathcal{W} \subseteq \mathfrak{V}(20\ell_2 \log n)$.

We regard the conditional expectation in Equation 2.14 as a random variable. The interpretation is that the expected number of a_ℓ 's direct children is less than $\log^{-3} n$, regardless of the global configuration at the infection time of a_ℓ , as long as it lies in $\mathfrak{V}(2\ell_2 \log^{-4} n)$ at that time.

Proof. We focus on the case when $S_0^1 \in \mathcal{W}$; otherwise Tr' is empty and the lemma trivially holds. First observe that all $a_j \in \text{Isd}_{T(\ell)}(S_{T(\ell)}^\ell, \ell_1 \log^{-1} n)$ are infected at or before the time a_ℓ is infected. Therefore they cannot be direct children of a_ℓ by Definition 2.3.9 and 2.3.10. On the other hand, an agent a_j is outside $\text{Isd}_{T(\ell)}(S_{T(\ell)}^\ell, \ell_1 \log^{-1} n)$

only if it is outside the ball $\mathbb{B}(S_{T(\ell)}^\ell, \ell_1 \log^{-1} n)$. Hence $\text{dchild}'(\mathbf{a}_\ell)$ is bounded by the number of agents initially outside $\mathbb{B}(S_{T(\ell)}^\ell, \ell_1 \log^{-1} n)$ that meet \mathbf{a}_ℓ before time Δt .

We consider two cases:

Case 1. $D_{T(\ell)} = 1$. By Lemma 2.3.7, we have

$$\mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, D_{T(\ell)} = 1, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \leq \log^{-3} n.$$

Case 2. $D_{T(\ell)} = 0$. By Definition 2.3.10, we have

$$\mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, D_{T(\ell)} = 0, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] = 0 \leq \log^{-3} n.$$

Therefore,

$$\begin{aligned} & \mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n) \right] \\ &= \mathbb{E}_{\mathcal{F}_{T(\ell)}} \left[\mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n), D_{T(\ell)} \right] \right] \\ &\leq \log^{-3} n. \end{aligned}$$

□

Recursive utilization of Lemma 2.3.13 on successive tree levels leads to the following lemma:

Lemma 2.3.14. *Consider a diffusion process with the island diffusion rule starting with an infected island $\text{Isd}_0(\mathbf{a}_1, \ell_1 \log^{-1} n)$. For the stopped diffusion tree Tr' with respect to any $\mathcal{W} \subseteq \mathfrak{V}(20\ell_2 \log n)$, let $\text{Height}(\text{Tr}')$ be its height. Then we have*

$$\Pr[\text{Height}(\text{Tr}') > 2 \log n] \leq \exp(-3 \log n \log \log n). \quad (2.15)$$

Let us denote the set of agents at the k th level as \mathbb{F}_k . It is worth pointing out that, despite a similar analysis to that of standard branching process, there is a technical complication on the conditioning argument since the creation of each child within the *same* level can be performed at *different* times in the diffusion process. This implies that there is no single filtration that we can condition on each level to analyze the expected size of the next one. Nevertheless, conditioning can be tailored to *each* agent at the same level.

Proof. We focus on the case when Tr' is non-trivial i.e. $S_0^1 \in \mathcal{W}$. Let $I(A) = 1$ if A occurs and 0 otherwise. We have, for any $k < 2 \log n$,

$$\begin{aligned}
& \mathbb{E}[|\mathbb{F}'_{k+1}| | \mathcal{F}_0, S_0^1 \in \mathcal{W}] \\
&= \mathbb{E} \left[\sum_{\mathbf{a}_\ell \in \mathbb{F}'_k} \sum_{\mathbf{a}_{\ell'} \in \text{dchild}'(\mathbf{a}_\ell)} |\text{Isd}_{T(\ell')}(\mathbf{a}_{\ell'}, \ell_1 \log^{-1} n) I(G_{T(\ell')} = 1)| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right] \\
&\leq (3 \log n) \mathbb{E} \left[\sum_{\mathbf{a}_\ell \in \mathbb{F}'_k} |\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right]
\end{aligned}$$

The equality holds by the stopping rule and the inequality holds by the small islands

property. Next we have

$$\begin{aligned}
& \mathbb{E} \left[\sum_{\mathbf{a}_\ell \in \mathbb{F}'_k} |\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right] \\
&= \mathbb{E} \left[\sum_{\ell \in [m]} I(\mathbf{a}_\ell \in \mathbb{F}'_k) |\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right] \\
&= \sum_{\ell \in [m]} \mathbb{E} \left[I(\mathbf{a}_\ell \in \mathbb{F}'_k) |\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right] \\
&= \sum_{\ell \in [m]} \mathbb{E} \left[\mathbb{E} \left[I(\mathbf{a}_\ell \in \mathbb{F}'_k) |\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W} \right] \middle| \mathcal{F}_1, S_0^1 \in \mathcal{W} \right] \\
&= \sum_{\ell \in [m]} \mathbb{E} \left[I(\mathbf{a}_\ell \in \mathbb{F}'_k) \mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W} \right] \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right]
\end{aligned}$$

The last equality holds because $I(\mathbf{a}_\ell \in \mathbb{F}'_k)$ is $\mathcal{F}_{T(\ell)}$ -measurable. Note that $S_0^1 \in \mathcal{W} \subseteq \mathfrak{V}(20\ell_2 \log n)$ implies $S_{T(\ell)}^\ell \in (20\ell_2 \log n - 4k\ell_2) \subset \mathfrak{V}(2\ell_2 \log^{-4} n)$ if $G_{T(\ell)} = 1$, by using Lemma 2.3.12. Therefore, by Lemma 2.3.13

$$\begin{aligned}
& \mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W}, G_{T(\ell)} = 1 \right] \\
&= \mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_{T(\ell)}^\ell \in \mathfrak{V}(2\ell_2 \log^{-4} n), G_{T(\ell)} = 1 \right] \leq \log^{-3} n
\end{aligned}$$

On the other hand,

$$\mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W}, G_{T(\ell)} = 0 \right] = 0$$

by the stopping rule. This leads to

$$\mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W} \right] \leq \log^{-3} n$$

which implies

$$\begin{aligned}
& \sum_{\ell \in [m]} \mathbb{E} \left[I(\mathbf{a}_\ell \in \mathbb{F}'_k) \mathbb{E} \left[|\text{dchild}'(\mathbf{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S_0^1 \in \mathcal{W} \right] \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W} \right] \\
&\leq \log^{-3} n \mathbb{E}[|\mathbb{F}'_k| \middle| \mathcal{F}_0, S_0^1 \in \mathcal{W}]
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E}[|\mathbb{F}'_{k+1}| \mid \mathcal{F}_0, S_0^1 \in \mathcal{W}] \\
& \leq 3 \log^{-2} n \mathbb{E}[|\mathbb{F}'_k| \mid \mathcal{F}_0, S_0^1 \in \mathcal{W}] \\
& \leq (3 \log^{-2} n)^k \mathbb{E}[|\mathbb{F}'_1| \mid \mathcal{F}_0, S_0^1 \in \mathcal{W}] \\
& \leq \log n (3 \log^{-2} n)^k
\end{aligned}$$

and hence

$$\Pr[|\mathbb{F}'_{2 \log n}| > 0] \leq \mathbb{E}[|\mathbb{F}'_{2 \log n}|] \leq \exp(-3 \log n \log \log n).$$

by combining with the case $a_1 \notin \mathcal{W}$. \square

We now prove Proposition 2.3.3.

Proof of Proposition 2.3.3. First note that the set of infected agents in a diffusion process with island diffusion rule, namely Definition 2.3.8, is always a superset of the coupled original diffusion process using Definition 2.1.1, at any time from 0 to Δt .

Next we have

$$\begin{aligned}
& \Pr[\text{Height}(\text{Tr}) > 2 \log n] \\
& = \Pr[(\text{Height}(\text{Tr}) > 2 \log n) \wedge (\text{Height}(\text{Tr}) = \text{Height}(\text{Tr}'))] \\
& \quad + \Pr[(\text{Height}(\text{Tr}) > 2 \log n) \wedge (\text{Height}(\text{Tr}) \neq \text{Height}(\text{Tr}'))] \\
& \leq \Pr[\text{Height}(\text{Tr}') > 2 \log n] + \Pr[\text{Tr}' \neq \text{Tr}] \\
& \leq \exp(-3 \log n \log \log n) + \Pr[G = 1] \\
& \leq 2 \exp(-3 \log n \log \log n)
\end{aligned}$$

Therefore, we have

$$\Pr[(\text{Height}(\text{Tr}) \leq 2 \log n) \wedge (G = 1)] \geq 1 - 3 \exp(-3 \log n \log \log n).$$

We will show that the viruses can be covered by the ball $\mathbb{B}(S_0^1, 9\ell_2 \log n)$ when

$$(\text{Height}(\text{Tr}) \leq 2 \log n) \wedge (G = 1) \wedge (S_0^1 \in \mathcal{W}).$$

Fix arbitrary infected $a_\ell \in \mathbb{F}_k$ with $k \leq 2 \log n$. By Lemma 2.3.12, we have $\|S_0^1 - S_{T(\ell)}^\ell\|_1 \leq 8\ell_2 \log n$. Moreover, $G = 1$ implies that for all $0 \leq t' \leq \Delta t$, $\|S_{T(\ell)}^\ell - S_{t'}^\ell\|_1 \leq 3\ell_2 \log^{-4} n \leq \ell_2 \log n$. This suggests $\|S_{t'}^\ell - S_0^1\|_\infty \leq 9\ell_2 \log n$ for all $t' \in [0, \Delta t]$. Therefore, the virus does not escape the ball $\mathbb{B}(S_0^1, 9\ell_2 \log n)$ within time $[0, \Delta t]$. \square

2.3.2 From local to global process

This section will be devoted to proving Theorem 2.3.1 via Proposition 2.3.3, or in other words, to turn our local probabilistic bound into a global result on the diffusion time.

We note that Proposition 2.3.3 deals with the case when there is only one initially infected agent. As discussed briefly in the discussion following the proposition, we want to iterate this estimate so that at every time increment Δt , the infected region is constrained within a certain radius from the initial positions of all the agents that are already infected at the start of the increment. Our argument is aided by noting which agents infect other agents. To ease the notation for this purpose, we introduce an artificial concept of virus type, denoted by $\nu_{i,t}$. We say an agent gets a virus of type $\nu_{i,t}$ if the meeting events of this agent can be traced upstream to the agent a_i , where a_i is already infected at time t . In other words, assume that a_i is infected at time t , and imagine that we remove the viruses in all infected agents except a_i but we keep the same dynamics of all the random walks. We say a particular agent gets $\nu_{i,t}$ if it eventually gets infected under this imaginary scenario. Note that under this

artificial framework of virus types it is obvious that an agent can get many different types of virus, in terms of both i and t .

In parallel to Proposition 2.3.3, we introduce the family of binary random variables $b_{i,t}$ to represent whether a virus of type $\nu_{i,t}$ can be constrained in a ball with radius $9\ell_2 \log n$:

Definition 2.3.15 ($b_{i,t}$ and virus of type $\nu_{i,t}$). Let $\overline{\mathfrak{B}} = \mathbb{B}(P, \frac{n}{4})$ where $P = (n/2, n/2, n/2)$. Let a_1, \dots, a_m be agents that are uniformly distributed on \mathcal{V}^3 at $t = 0$ and diffuse according to Definition 2.1.1. Let t be an arbitrary time step and $i \in [m]$. At time t , a virus of type $\nu_{i,t}$ emerges on agent a_i and diffuses. Define the binary random variable $b_{i,t}$ as follows:

- If $S_t^i \in \overline{\mathfrak{B}}$: $b_{i,t}$ is set as 1 if and only if all the agents infected by the virus of type $\nu_{i,t}$ at time $t + \Delta t$ can be covered by the ball $\mathbb{B}(S_t^i, 9\ell_2 \log n)$.
- If $S_t^i \notin \overline{\mathfrak{B}}$: $b_{i,t} = 1$.

Let us start with showing $b_{i,t} = 1$ for all i and t with high probability:

Corollary 2.3.16. Consider the family of random variables $\{b_{i,t} : i \in [m], t \leq n^{2.5}\}$ defined above. We have

$$\Pr \left[\bigwedge_{i \in [m], t \leq n^{2.5}} (b_{i,t} = 1) \right] \geq 1 - \exp(-4 \log n \log \log n).$$

Proof. We first bound $\Pr[b_{i,t} = 1]$ for any specific i and t . Since the agents are placed according to stationary distribution at $t = 0$, each agent is still distributed uniformly at time t . Next, at time t , we may relabel the agents so that a_i is regarded as the

single initially infected agent in Proposition 2.3.3, where \mathcal{W} is set as $\overline{\mathfrak{B}}$. We therefore have $\Pr[b_{i,t} = 1] \geq 1 - \exp(-5 \log n \log \log n)$.

Next, we may apply a union bound across all i and t to get the desired result. \square

Lemma 2.3.17. *Let $\mathfrak{B} = \mathbb{B}(P, n/8)$. Let B_t be the indicator variable that there is at least one agent in \mathfrak{B} at time t . Let $B = \prod_{t \leq n^{2.5}} B_t$, the indicator variable that there is at least one agent in \mathfrak{B} at all times in $[0, n^{2.5}]$. We have*

$$\Pr[B = 0] \leq \exp(-\log^2 n)$$

for sufficiently large n .

Proof. First, notice that for any specific t , the expected number of agents in \mathfrak{B} is $\Omega(m)$. Therefore, by Chernoff bound (using the version in Theorem A.1.1) $\Pr[B_t = 0] \leq \exp(-\Omega(m)) \leq \exp(-\log^3 n)$. Next, by a union bound, we have $\Pr[B = 0] \leq n^{2.5} \exp(-\log^3 n) \leq \exp(-\log^2 n)$. \square

We next present our major lemma for this subsection.

Lemma 2.3.18. *Let a_1, \dots, a_m be placed uniformly at random on \mathcal{V}^3 such that $m \geq 1600n \log^2 n$. Let $\ell_2 = \sqrt{n^3/m}$. Let $\{b_{i,t} : i \in [m], t \leq n^{2.5}\}$ and B be the random variables described above. If $b_{i,t} = 1$ for all i, t and $B = 1$, then the diffusion time is at least $T_c = \frac{1}{81} \ell_2 n \log^{-29} n$.*

Notice that by Proposition 2.3.3 and Lemma 2.3.17,

$$\begin{aligned} \Pr \left[\bigwedge_{i \leq m, t \leq n^{2.5}} (b_{i,t} = 1) \right] &\geq 1 - \exp(-4 \log n \log \log n). \\ \Pr[B = 1] &\geq 1 - \exp(-\log^2 n). \end{aligned}$$

Together with Lemma 2.3.18, Theorem 2.3.1 then follows.

Proof. Without loss of generality, we assume the x , y , and z coordinates of S_0^1 are all negative. We can always rotate the space \mathcal{V}^3 at $t = 0$ correspondingly to ensure this assumption to hold.

We shall prove by contradiction. Consider two balls \mathfrak{B} and $\overline{\mathfrak{B}}$ defined above. Assume the diffusion time is less than T_c . First, because $B = 1$, a necessary condition for the diffusion to complete is that an infected agent ever visits the smaller ball \mathfrak{B} at a time $T' \leq T_c$ (since otherwise the agents in \mathfrak{B} would be uninfected all the time, including at T_c). We call this agent $a_{i'}$. Next, for the infection to get into \mathfrak{B} , it must happen that there is an infected agent that enters $\overline{\mathfrak{B}}$ from outside, whose infection trajectory eventually reaches $a_{i'}$. We denote T'' to be the *last* time that this happens, and the responsible agent to be $a_{i''}$. We focus on the trajectory of infection that goes from $a_{i''}$ to $a_{i'}$ that lies completely inside $\overline{\mathfrak{B}}$ (which exists since T'' is the last time of entry). Note that we consider at most $\lceil T_c/\Delta t \rceil$ time increments of Δt . Now, since $b_{i,t} = 1$ for all i and t , by repeated use of triangle inequality, we get

$$\begin{aligned}
 \|S_{T'}^{i'} - S_{T''}^{i''}\|_{\infty} &\leq 9\ell_2 \log n \left\lceil \frac{T_c}{\Delta t} \right\rceil \\
 &\leq 9\ell_2 \log n \left(\frac{(1/81)\ell_2 n \log^{-29} n}{\ell_2^2 \log^{-28} n} + 1 \right) \\
 &\leq \frac{n}{9} + 9\ell_2 \log n \\
 &< \frac{n}{8} - 1.
 \end{aligned}$$

On the other hand, the physical dimensions of \mathfrak{B} and $\overline{\mathfrak{B}}$ give that

$$\|S_{T'}^{i'} - S_{T''}^{i''}\|_{\infty} \geq \frac{n}{8} - 1$$

which gives a contradiction. □

2.4 Upper bound

We now focus on an upper bound for the diffusion time. Our main result is the following:

Theorem 2.4.1. *Let a_1, \dots, a_m be placed uniformly at random on \mathcal{V}^3 , where $n \leq m \leq n^3$. Let $\hat{\ell}_2 = \sqrt{n^3/m} \cdot \log n$. When n is sufficiently large, the diffusion time T satisfies*

$$\Pr[T \geq 128n\hat{\ell}_2 \log^{47} n] \leq \exp(-\frac{1}{2} \log^2 n).$$

Note that this theorem shows that an upper bound of $\tilde{O}(n\sqrt{n^3/m})$ holds for the diffusion time with high probability. Hence the upper and lower bounds “match” up to logarithmic factors. We remark that the constant 47 in the exponent has not been optimized.

The main goal of this section is to prove this theorem. Our proof strategy relies on calculating the growth rate of the *total* infected agents evolving over time; such growth rate turns out to be best characterized as the increase/decrease in infected/uninfected agents relative to the size of the corresponding population. More precisely, we show that for a well-chosen time increment, either the number of infected agents doubles or the number of uninfected agents reduces by half with high probability. The choice of time increment is complex, depending on the analysis of the local interactions in small cubes and the global geometric arrangements of these cubes with respect to the distribution of infected agents.

As with the lower bound proof, our technique for proving Theorem 2.4.1 is different from existing methods. Let us briefly describe them and explain the challenges in extending to higher dimensional cases; further details are left to Chapter 2.7.2.

Roughly, existing methods can be decomposed into two steps (see for example [92]):

- 1) In the first step, consider a small ball of length r that contains the initially infected agent. One can see that for $d = 2$, when number of agents in the ball is $\tilde{\Theta}(m(r/n)^2)$, within time increment r^2 the number of infections to agents initially in this ball is $\tilde{\Omega}(1)$ w.h.p..
- 2) The 2nd step is to prove that for any ball that has $\tilde{\Omega}(1)$ infected agents at time t , its surrounding adjacent balls will also have $\tilde{\Omega}(1)$ infected agents by time $t + r^2$. From these two steps, one can recursively estimate the time to spread infection across the whole space \mathcal{V}^2 to be $n/r \times r^2 = nr$ w.h.p.. In other words, at time nr all the balls in \mathcal{V}^2 will have $\tilde{\Omega}(1)$ infected agents. Moreover, every agent in \mathcal{V}^2 is infected in the same order of time units, because $\tilde{\Omega}(1)$ is also the total number of agents in any ball under good density condition. Finally, it is then clear that a good choice of r is then n/\sqrt{m} , which would give the optimal upper bound.

The critical difference in the analysis for $d > 2$ lies primarily in the magnitude of the meeting probability of random walks. In the case of $d = 2$, the meeting probability of two random walks at distance r within time r^2 is $\tilde{\Theta}(1)$, whereas for $d > 2$ the meeting probability is $\Theta(1/r^{d-2})$. For $d = 2$, this means that it is easy, i.e. w.h.p., for infection to transmit from a ball with $\tilde{\Omega}(1)$ infected agents to an adjacent uninfected ball, so that the latter also has $\tilde{\Omega}(1)$ infected agents after a time increment of r^2 . In the case $d > 2$, however, $\Omega(r^{d-2})$ infected agents must be present in a ball to transmit virus effectively to its adjacent uninfected ball within r^2 time. Consequently, arguing for transmission across adjacent balls becomes problematic (more details are in Chapter 2.7.2). In light of this, we take an alternate approach to analyze both the local interactions and the global distribution of infected agents. Instead of focusing on

transmission from one infected ball to another, we calculate the spreading rate across the whole space. This turns out to be fruitful in obtaining a tight upper bound.

We briefly describe the forthcoming analysis. As with the lower bound, we start with local analysis. We partition the space \mathcal{V}^3 into disjoint subcubes each of size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$. Here $\hat{\ell}_2$ is just a logarithmic factor larger than ℓ_2 , the size of subcubes used for the lower bound, so that with overwhelming probability there are *at least* $\hat{\ell}_2$ agents in a subcube. We show that, within every subcube, over a time increment of length $\Theta(\hat{\ell}_2^2)$ the number of infections is roughly a $\tilde{\Omega}(1)$ -fraction of the minimum of the number of infected and uninfected agents. Hence, at least locally, we have the desired behavior described above.

We then leverage the local analysis to obtain the global result. However, this is not straightforward. For example, consider the beginning when the number of infected agents is small. If infected agents are distributed uniformly throughout the whole space, it would be easy to show that new infections would roughly grow in proportion to the number of infected agents. However, if infected agents are concentrated into a small number of subcubes, we have to show that there are enough neighboring subcubes on the boundary of these infected subcubes that these subcubes become infected suitably rapidly, so that after the appropriate time increment the number of infected agents doubles. Similar arguments arise for the case when infected agents are dominant, with the end result being a halving of the uninfected population.

We now make the above discussion rigorous. First, let $b = (2n + 1)/\hat{\ell}_2$, so there are in total b^3 subcubes. As in the previous section, we divide the time into small intervals. We reuse the symbol Δt to represent the length of each interval but here we

set $\Delta t = 16\hat{\ell}_2^2$. Our local bound is built within each subcube (and pair of neighboring subcubes) in the time increment Δt :

Lemma 2.4.2. *Let $\mathcal{W} \subset \mathcal{V}^3$ be a region that can be covered by a ball of radius $2\hat{\ell}_2$ under the L_∞ -norm. Let A^f and A^u be subsets of infected and uninfected agents in \mathcal{W} at time t such that $|A^f| = m_1$, $|A^u| = m_2$, and $\max\{m_1, m_2\} = \hat{\ell}_2 / \log^2 n$. Given any initial placement of the agents of A^f and A^u , let $M(t)$ be the number of agents in A^u that become infected at time $t + \Delta t$. We have*

$$\Pr \left[M(t) \geq \frac{\tau_0 \min\{m_1, m_2\}}{\log^4 n} \middle| \mathcal{F}_t \right] \geq \tau_0 \log^{-6} n.$$

for some constant τ_0 , where \mathcal{F}_t denotes the information of the whole diffusion process up to time t .

Proof. The high level idea of our proof is to count the total number of times the infected agents meet the uninfected ones between time t and $t + \Delta t$. The probability two agents in \mathcal{W} can meet each other within time Δt is approximately $\tilde{\Omega}(1/\hat{\ell}_2)$ (Lemma 2.2.3). The expected number of meetings is thus $\tilde{\Omega}(1/\hat{\ell}_2) \times m_1 m_2 = \tilde{\Omega}(\min\{m_1, m_2\})$. The total number of newly infected agents is the number of meetings modulo possible overcounts on each originally uninfected agent. If we can show that the number of meetings is $\tilde{O}(1)$ for each uninfected agent, then we can conclude that $\tilde{\Omega}(\min\{m_1, m_2\})$ more agents become infected at time $t + \Delta t$.

Two problems need to be addressed to implement this idea. First, when the agents are close to the boundary, they may behave in a more complicated way than suggested by Lemma 2.2.3. Second, in the (rare) case that an uninfected agent is surrounded by a large number of infected agents, it can possibly meet with $\omega(1)$ infected agents, making it difficult to give an upper bound over the number of overcounts.

To address both problems, we wait $\hat{\ell}_2^2$ time steps before starting our analysis on infections. This time gap is enough to guarantee that with constant probability, the agents are “locally shuffled” so that by time $t + \hat{\ell}_2^2$,

1. all agents are reasonably far away from the boundaries,
2. the distance between any pair of agents is “appropriate” (in our case, the distance is between $\hat{\ell}_2$ and $9\hat{\ell}_2$).

Intuitively, the “local shuffling” works because central limit theorem implies that the agents’ distribution at the end of these steps is approximately multivariate Gaussian.

We now implement this idea. First, we couple the (sub)process in \mathcal{W} with one that has slower diffusion rule. In the coupled process, we first wait for $\hat{\ell}_2^2$ time steps, in which no agent becomes infected even if it meets an infected agent. After these $\hat{\ell}_2^2$ steps, for an arbitrary $a_i \in A^f$ and $a_j \in A^u$, let $X_{i,j} = 1$ if both of the following conditions hold,

- the L_1 -distance between a_i and a_j is between $\hat{\ell}_2$ and $9\hat{\ell}_2$.
- the L_1 -distance between a_i and any boundary is at least $360\hat{\ell}_2$.

By Corollary 2.6.9, $\Pr[X_{i,j} = 1] \geq \tau$ for some constant τ . Therefore,

$$\mathbb{E}\left[\sum_{a_i \in A^f, a_j \in A^u} X_{i,j}\right] \geq \tau m_1 m_2.$$

On the other hand, $\sum_{i,j} X_{i,j} \leq m_1 m_2$. It follows easily that we have $\Pr[\sum_{i,j} X_{i,j} \geq \frac{1}{2}\tau m_1 m_2] \geq \tau/2$.

Our slower diffusion rule then allows $a_i \in A^f$ to transmit its virus to $a_j \in A^u$ if and only if

- $X_{i,j} = 1$,
- they meet during the time interval $(t + \hat{\ell}_2^2, t + \Delta t]$,
- a_i and a_j have not visited any boundary after $t + \hat{\ell}_2^2$ before they meet. In other words, an agent $a_i \in A^f$ ($a_j \in A^u$ resp.) loses its ability to transmit (receive resp.) the virus when it hits a boundary after the initial waiting stage.

An added rule is that agents in A^u will not have the ability to transmit the virus even after they are infected.

Let $Y_{i,j}$ be the indicator random variable that is set to 1 if and only if a_i transmits its virus to a_j under the slower diffusion rule. By Lemma 2.2.6, we have $\Pr[Y_{i,j} = 1 \mid X_{i,j} = 1] = \Omega(1/\hat{\ell}_2)$. Therefore, we have

$$\Pr[Y_{i,j} = 1] \geq \Pr[Y_{i,j} = 1 \mid X_{i,j} = 1] \Pr[X_{i,j} = 1] = \Omega(1/\hat{\ell}_2).$$

Hence,

$$\mathbb{E}\left[\sum_{a_i \in A^f, a_j \in A^u} Y_{i,j}\right] = \Omega(m_1 m_2 / \hat{\ell}_2) \geq \tau_1 m_1 m_2 / \hat{\ell}_2$$

for some constant τ_1 . $\sum_{a_i \in A^f, a_j \in A^u} Y_{i,j}$ is approximately the number of newly infected agents except that the same agent in A^u may be counted multiple times. Our next task is thus to give an upper bound on the number of overcounts. Specifically, we fix an agent $a_j \in A^u$ and argue that the probability $\sum_{a_i \in A^f} Y_{i,j} \geq \log^2 n$ is small.

In our slower diffusion model, once an agent in A^f reaches the boundary, it is not able to transmit the virus further. We need to bound the probability that there

are more than $\log^2 n$ agents in A^f that transmit the virus to a_j *before they hit any boundary*. This probability is at most the probability that more than $\log^2 n$ infected agents performing *unbounded* random walks meet a_j , where each infected agent is at least $\hat{\ell}_2$ away from a_j initially.

By Lemma 2.2.4, there exists a constant c_0 such that for all possible values of $X_{i,j}$ and sufficiently large n :

$$\begin{aligned}
& \Pr\left[\sum_{a_i \in A^f} Y_{i,j} \geq \log^2 n \mid X_{1,j}, X_{2,j}, \dots, X_{m_1,j}\right] \\
& \leq \left(\frac{\sum_{i \leq m_1} X_{i,j}}{\log^2 n}\right) \left(\frac{c_0 \log^2 n}{\hat{\ell}_2}\right)^{\log^2 n} \\
& \leq \left(\frac{m_1}{\log^2 n}\right) \left(\frac{c_0 \log^2 n}{\hat{\ell}_2}\right)^{\log^2 n} \\
& \leq \left(\frac{em_1}{\log^2 n}\right)^{\log^2 n} \left(\frac{c_0 \log^2 n}{\hat{\ell}_2}\right)^{\log^2 n} \\
& = \left(\frac{c_0 em_1}{\hat{\ell}_2}\right)^{\log^2 n} \\
& \leq \exp(-\log^2 n \log \log n).
\end{aligned}$$

Therefore,

$$\Pr\left[\sum_{a_i \in A^f} Y_{i,j} \geq \log^2 n\right] = \mathbb{E}\left[\Pr\left[\sum_{a_i \in A^f} Y_{i,j} \geq \log^2 n \mid X_{1,j}, \dots, X_{m_1,j}\right]\right] \leq \exp(-\log^2 n \log \log n).$$

By a union bound we have

$$\Pr[\exists j : \sum_{a_i \in A^f} Y_{i,j} \geq \log^2 n] \leq m \cdot \exp(-\log^2 n \log \log n) \leq \exp(-2 \log^2 n). \quad (2.16)$$

Next, let us fix $a_i \in A^f$ and we may argue in a similar way to obtain

$$\Pr[\exists i : \sum_{a_j \in A^u} Y_{i,j} \geq \log^2 n] \leq \exp(-2 \log^2 n).$$

Define e_t as the event that $(\forall i, \sum_{a_j \in A^u} Y_{i,j} \leq \log^2 n) \wedge (\forall j, \sum_{a_i \in A^f} Y_{i,j} \leq \log^2 n)$. Therefore, $\Pr[e_t] \geq 1 - 2 \exp(-2 \log^2 n)$. Observe that e_t implies $\sum_{i,j} Y_{i,j} \leq \min\{m_1, m_2\} \log^2 n$. We have

$$\begin{aligned}
\tau_1 m_1 m_2 / \hat{\ell}_2 &\leq \mathbb{E}[\sum_{i,j} Y_{i,j}] \\
&= \mathbb{E}[\sum_{i,j} Y_{i,j} | e_t] \Pr[e_t] + \mathbb{E}[\sum_{i,j} Y_{i,j} | \neg e_t] \Pr[\neg e_t] \\
&\leq \mathbb{E}[\sum_{i,j} Y_{i,j} | e_t] + m^2 \Pr[\neg e_t] \\
&\leq \mathbb{E}[\sum_{i,j} Y_{i,j} | e_t] + 2m^2 \exp(-2 \log^2 n).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\sum_{i,j} Y_{i,j} | e_t] &\geq \tau_1 m_1 m_2 / \hat{\ell}_2 - 2m^2 \exp(-2 \log^2 n) \\
&\geq \frac{\tau_1 m_1 m_2}{2\hat{\ell}_2} \\
&= \frac{\tau_1 \max\{m_1, m_2\} \min\{m_1, m_2\}}{2\hat{\ell}_2} \\
&\geq \frac{\tau_1 \min\{m_1, m_2\}}{2 \log^2 n}
\end{aligned}$$

Next, define indicator variable $I_j = 1$ if and only if $\sum_{a_i \in A^f} Y_{i,j} > 0$. The sum $\sum_j I_j$ is the total number of newly infected agents in our weaker process and thus is a lower bound on $M(t)$. Note that if e_t holds,

$$\sum_{a_j \in A^u} I_j \leq \sum_{i,j} Y_{i,j} \leq \min\{m_1, m_2\} \log^2 n,$$

and hence

$$\mathbb{E}[\sum_{a_j \in A^u} I_j | e_t] \leq \min\{m_1, m_2\} \log^2 n.$$

On the other hand,

$$\mathbb{E}[\sum_j I_j | e_t] \geq \log^{-2} n \mathbb{E}[\sum_{i,j} Y_{i,j} | e_t] \geq \frac{\tau_1 \min\{m_1, m_2\}}{2 \log^4 n} \quad (2.17)$$

Now define $\tilde{m} = \frac{\tau_1 \min\{m_1, m_2\}}{4 \log^4 n}$. We have

$$\begin{aligned}
& 2\tilde{m} \\
& \leq \mathbb{E}\left[\sum_j I_j \mid e_t\right] \\
& = \mathbb{E}\left[\sum_j I_j \mid e_t, \sum_j I_j \leq \tilde{m}\right] \Pr\left[\sum_j I_j \leq \tilde{m} \mid e_t\right] \\
& \quad + \mathbb{E}\left[\sum_j I_j \mid e_t, \sum_j I_j > \tilde{m}\right] \Pr\left[\sum_j I_j > \tilde{m} \mid e_t\right] \\
& \leq \mathbb{E}\left[\sum_j I_j \mid e_t, \sum_j I_j \leq \tilde{m}\right] + \min\{m_1, m_2\} \log^2 n \Pr\left[\sum_j I_j > \tilde{m} \mid e_t\right] \\
& \leq \tilde{m} + \min\{m_1, m_2\} \log^2 n \Pr\left[\sum_j I_j > \tilde{m} \mid e_t\right].
\end{aligned}$$

Therefore,

$$\Pr\left[\sum_j I_j > \frac{\tau_1 \min\{m_1, m_2\}}{4 \log^4 n} \mid e_t\right] \geq \frac{\tau_1}{4 \log^6 n}.$$

Finally,

$$\begin{aligned}
\Pr\left[\sum_j I_j > \frac{\tau_1 \min\{m_1, m_2\}}{4 \log^4 n}\right] & \geq \Pr\left[\sum_j I_j > \frac{\tau_1 \min\{m_1, m_2\}}{4 \log^4 n} \mid e_t\right] \Pr[e_t] \\
& \geq \frac{\tau_1}{4 \log^6 n} (1 - 2 \exp(-\log^2 n)) \geq \frac{\tau_1}{5 \log^6 n}
\end{aligned}$$

By setting $\tau_0 = \tau_1/5$, we get our result. □

The next step is to characterize the growth rate at a larger scale. This requires more notation. We denote the set of b^3 subcubes of size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$ as $\mathfrak{C} = \{h_{i,j,k} : i, j, k \in [b]\}$. For an arbitrary subcube $h_{i,j,k}$, we define its neighbors as $N(h_{i,j,k}) = \{h_{i',j',k'} : |i - i'| + |j - j'| + |k - k'| = 1\}$. In other words, $h_{i',j',k'}$ is a neighbor of $h_{i,j,k}$ if and only if both subcubes share a facet. Let \mathcal{H} be an arbitrary subset of \mathfrak{C} . We write $N(\mathcal{H}) = \bigcup_{h \in \mathcal{H}} N(h)$.

Definition 2.4.3 (Exterior and interior surface). *Let \mathcal{H} be a subset of \mathfrak{C} . The exterior surface of \mathcal{H} is $\partial\mathcal{H} = N(\mathcal{H}) - \mathcal{H}$. Let $\overline{\mathcal{H}}$ be the complement of \mathcal{H} . The interior surface of \mathcal{H} is $\dot{\partial}\mathcal{H} = N(\overline{\mathcal{H}}) - \overline{\mathcal{H}}$, i.e., the exterior surface of the complement of \mathcal{H} .*

At time step $t = i\Delta t$, let \mathcal{G}_t be the set of all subcubes that contain more than $\hat{\ell}_2/2$ infected agents and let $g_t = |\mathcal{G}_t|$; let $\mathcal{B}_t = \overline{\mathcal{G}_t}$ be the rest of the subcubes and let $b_t = |\mathcal{B}_t|$. We say a subcube in \mathcal{G}_t an *infected* (*good*) subcube and a subcube in \mathcal{B}_t an *uninfected* (*bad*) subcube.

We classify the agents in the process according to the subcubes they reside in. To facilitate our analysis, we adopt the notational system \mathfrak{A}_t and \mathfrak{A}_t^f to represent the total number of agents that belong to the type specified in the superscript. Specifically, let \mathfrak{A}_t^f be the set of infected agents at time t ; decompose the set \mathfrak{A}_t^f as $\mathfrak{A}_t^f = \mathfrak{A}_t^{f,\mathcal{G}} \cup \mathfrak{A}_t^{f,\mathcal{B}}$, where $\mathfrak{A}_t^{f,\mathcal{G}}$ is the set of infected agents residing in the subcubes in \mathcal{G}_t and $\mathfrak{A}_t^{f,\mathcal{B}}$ the set of infected agents in \mathcal{B}_t . Similarly, let \mathfrak{A}_t^u be the set of all uninfected agents; decompose the set \mathfrak{A}_t^u as $\mathfrak{A}_t^u = \mathfrak{A}_t^{u,\mathcal{G}} \cup \mathfrak{A}_t^{u,\mathcal{B}}$, where $\mathfrak{A}_t^{u,\mathcal{G}}$ is the set of uninfected agents residing in the subcubes in \mathcal{G}_t and $\mathfrak{A}_t^{u,\mathcal{B}}$ the set of uninfected agents in \mathcal{B}_t . Furthermore, we denote $\Delta\mathfrak{A}_t^{\mathcal{G}}$ and $\Delta\mathfrak{A}_t^{\mathcal{B}}$ as the set of agents in \mathcal{G}_t and \mathcal{B}_t respectively that are infected between t and $t + \Delta t$. Hence the total increase in infected agents, or equivalently the total decrease in uninfected agents, between t and $t + \Delta t$ is given by $\Delta\mathfrak{A}_t = \Delta\mathfrak{A}_t^{\mathcal{G}} \cup \Delta\mathfrak{A}_t^{\mathcal{B}}$. Lastly, we let $\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}}$ be the set of agents in $\mathcal{G}_t \cup \partial\mathcal{G}_t$ that are infected between t and $t + \Delta t$.

Similar to the lower bound analysis, here we also introduce good density conditions that can be easily verified to hold with high probability, and reuse the symbols D_t and D with slightly different meanings from the last section:

Definition 2.4.4. Let $\{D_t : t \geq 0\}$ be a sequence of binary random variables such that $D_t = 1$ if for all time steps on or before t , the number of agents for any subcube in \mathcal{V}^3 with size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$ is between $\hat{\ell}_2$ and $2\hat{\ell}_2 \log^2 n$. Also, let $D = D_{n^{2.5}}$.

The following lemma shows that $D_t = 1$ with high probability:

Lemma 2.4.5. For any $t \leq n^{2.5}$, $\Pr[D_t = 0] \leq \exp(-\frac{1}{15} \log^2 n)$ for sufficiently large n .

Proof of Lemma 2.4.5. Fix a time t and let \tilde{m} be the number of agents in an arbitrary subcube of size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$. We have $\mathbb{E}[\tilde{m}] \geq \hat{\ell}_2 \log^2 n / 27 \geq \log^2 n$. Therefore, by Chernoff bounds (Theorem A.1.1), $\Pr[\tilde{m} \in [\frac{1}{2}\mathbb{E}[\tilde{m}], \frac{3}{2}\mathbb{E}[\tilde{m}]]] \geq 1 - 2\exp(-\log^2 n / 12)$. Now the total number of possible subcubes is at most $(2n+1)^3$ and the total number of time steps is $n^{2.5}$. By a union bound, we have

$$\Pr[D = 0] \leq (2n+1)^3 \cdot n^{2.5} \cdot 2\exp(-\frac{1}{12} \log^2 n) \leq \exp(-\frac{1}{15} \log^2 n)$$

for sufficiently large n . □

We now state two bounds on the growth rate of the agent types, one relative to the “boundary subcubes” $\partial\mathcal{G}_t$ and one relative to the total agents of each type:

Corollary 2.4.6. For some constant τ_0 ,

$$\Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

Consequently,

$$\Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n$$

and

$$\Pr \left[|\Delta\mathfrak{A}_t^{\mathcal{B}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

Corollary 2.4.7. *We have*

$$\Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \geq \frac{\tau_0^2}{4 \log^{38} n} |\mathfrak{A}_t^{u, \mathcal{G}}| \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n$$

and

$$\Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \geq \frac{\tau_0^2}{4 \log^{38} n} |\mathfrak{A}_t^{f, \mathcal{B}}| \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

The proofs of these two corollaries both rely on using coupled diffusion processes that have slower diffusion rates. These processes only allow infection locally i.e. within each “pair” of subcubes on the surface of \mathcal{G}_t in the case of Corollary 2.4.6 and within each subcube in Corollary 2.4.7, and hence can be tackled by Lemma 2.4.2. The surface $\partial \mathcal{G}_t$ in Corollary 2.4.6 appears naturally from a matching argument between neighboring infected and uninfected subcubes. Roughly speaking, the bounds in Corollary 2.4.6 are tighter and hence more useful for the cases where infected/uninfected agents are dense in the infected/uninfected subcubes, while those in Corollary 2.4.7 are for cases where the agent types are more uniformly distributed.

Proof of Corollary 2.4.6. Agents in $\widetilde{\Delta \mathfrak{A}_t^{\mathcal{G}}} \cap \Delta \mathfrak{A}_t^{\mathcal{B}}$ are those initially in $\partial \mathcal{G}_t$ at t and become infected at the time $t + \Delta t$. We focus on how the uninfected agents in $\partial \mathcal{G}_t$ become infected.

Let us construct a graph $G = (V, E)$, in which the vertex set V of G consists of subcubes in $\partial \mathcal{G}_t \cup \dot{\partial} \mathcal{G}_t$ and the edge set is defined as

$$E = \{ \{u, v\} : u \in \partial \mathcal{G}_t, v \in \dot{\partial} \mathcal{G}_t, u \in N(v) \}.$$

We may use a greedy algorithm to argue that there is a matching on G from $\partial \mathcal{G}_t$ to $\dot{\partial} \mathcal{G}_t$ with size at least $|\partial \mathcal{G}_t|/11$. Specifically, we have the following lemma.

Lemma 2.4.8. *Let \mathcal{G} be an arbitrary subset of \mathfrak{C}^d . Define $\dot{\partial}\mathcal{G}$ and $\partial\mathcal{G}$ as the interior and exterior surfaces of \mathcal{G} (i.e. the set of points in \mathcal{G} that neighbor with \mathcal{G}^c and the set of points in \mathcal{G}^c that neighbor with \mathcal{G} resp.; \vec{u} and \vec{v} are neighbors if $\|\vec{u} - \vec{v}\|_1 = 1$). Define a bipartite graph with nodes denoting $\dot{\partial}\mathcal{G}$ and $\partial\mathcal{G}$, in which an edge $(\vec{u}, \vec{v}), \vec{u} \in \dot{\partial}\mathcal{G}, \vec{v} \in \partial\mathcal{G}$ exists whenever \vec{u} and \vec{v} are neighbors. Then there exists a matching M in this graph with $|M| \geq |\partial\mathcal{G}|/(4d - 1)$.*

Proof of Lemma 2.4.8. We prove this statement by explicitly constructing the matching M . First notice that the degree of each node is in the range $[1, 2d]$. We build M iteratively. Each time, we pick an edge $(\vec{u}, \vec{v}) \in E$ and place the edge into M . We then remove nodes \vec{u}, \vec{v} from L and R respectively as well as all edges incident to them. Since the degrees of \vec{u}, \vec{v} are bounded by $2d$, we will remove at most $4d - 1$ edges from E . We continue this process until no edge is left. Clearly, the edges we place into M form a matching. Because there are at least $|\partial\mathcal{G}|$ number of edges by the lower bound of degrees, we conclude that $|M| \geq \frac{|\partial\mathcal{G}|}{4d-1}$. \square

Denote the matching as

$$\mathfrak{M} = \{\{h_1, h'_1\}, \{h_2, h'_2\}, \dots, \{h_k, h'_k\} : h_i \in \partial\mathcal{G}_t, h'_i \in \dot{\partial}\mathcal{G}_t\},$$

where $k \geq |\partial\mathcal{G}_t|/11$. We next define a coupling process with a slower diffusion rule: an infected agent can transmit virus to an uninfected one if and only if at time t the infected agent is in h'_j and the uninfected one is in h_j for some j . Let ρ_j be the number of uninfected agents initially in h_j at time t that become infected by time $t + \Delta t$ under the slower diffusion rule. We have $\sum_{j \leq k} \rho_j$ at most $|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}|$ in the original process. We design the coupling in this way because ρ_j s are independent of

each other as they are decided by independent walks from disjoint pairs of subcubes.

Next we apply Lemma 2.4.2 on each pair of the matching. Fix an arbitrary matched pair $\{h_j, h'_j\}$. Since $h_j \in \partial\mathcal{G}_t$, at time t there are at least $\hat{\ell}_2/2$ uninfected agents in h_j ; similarly, since $h'_j \in \partial\mathcal{G}_t$, there are at least $\hat{\ell}_2/2$ infected agents in h'_j . At time t we can find a subset of uninfected agent A^u in h'_j and a subset of infected agents A^f in h_j such that $|A^u| = |A^f| = \hat{\ell}_2/\log^2 n$. Therefore, by Lemma 2.4.2, we have

$$\Pr[\rho_j \geq \frac{\tau_0 \hat{\ell}_2}{\log^4 n} \mid \mathcal{F}_t, D_t = 1] \geq \tau_0 \log^{-6} n \quad (2.18)$$

for some constant τ_0 . From Equation 2.18, we can see that $\mathbb{E}[\rho_j \mid \mathcal{F}_t, D_t = 1] \geq \tau_0^2 \hat{\ell}_2 \log^{-10} n$. Therefore,

$$\mathbb{E}[\sum_{j \leq k} \rho_j \mid \mathcal{F}_t, D_t = 1] \geq \frac{\tau_0^2}{11} |\partial\mathcal{G}_t| \hat{\ell}_2 \log^{-10} n. \quad (2.19)$$

Next, we consider two cases.

Case 1., $|\partial\mathcal{G}_t| \leq \log^9 n$. In this case

$$\begin{aligned} & \Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \mid \mathcal{F}_t, D_t = 1 \right] \\ & \geq \Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \geq \log^9 n \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \mid \mathcal{F}_t, D_t = 1 \right] \\ & \geq \Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \geq \frac{\tau_0 \hat{\ell}_2}{4 \log^4 n} \mid \mathcal{F}_t, D_t = 1 \right] \\ & \geq \Pr[\rho_1 \geq \tau_0 \hat{\ell}_2 \log^{-4} n \mid \mathcal{F}_t, D_t = 1] \end{aligned}$$

(only focus on an arbitrary matched pair in the matching)

$$\geq \tau_0 \log^{-6} n. \quad (\text{Lemma 2.4.2})$$

Case 2. $|\partial\mathcal{G}_t| > \log^9 n$. Notice that ρ_1, \dots, ρ_k are independent by construction. Also,

we have

$$|\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \leq \frac{\tau_0^2}{22} |\partial\mathcal{G}_t| \hat{\ell}_2 \log^{-10} n \leq \frac{1}{2} \mathbb{E} \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \middle| \mathcal{F}_t, D_t = 1 \right].$$

By a Chernoff bound (the version we use is Theorem A.1.1 with $\delta = 1/2$), we have

$$\begin{aligned} & \Pr \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \leq |\partial\mathcal{G}_t| \cdot \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t, D_t = 1 \right] \\ & \leq 2 \exp \left(- \frac{(\frac{1}{2})^2 \mathbb{E} \left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \middle| \mathcal{F}_t, D_t = 1 \right]}{3} \right) \\ & \leq 2 \exp \left(- \frac{\tau_0^2}{132} |\partial\mathcal{G}_t| \hat{\ell}_2 \log^{-10} n \right) \quad (\text{using Equation 2.19}) \\ & \leq 1 - \tau_0 \log^{-6} n \end{aligned}$$

for sufficiently large n . Our corollary thus follows. \square

Proof of Corollary 2.4.7. Let us start with proving the first inequality. We first couple the diffusion problem with a slower diffusion process defined as follows. First, all the infected agents in \mathcal{B}_t at time t cannot transmit the virus. Second, agents in \mathcal{G}_t are able to transmit the virus to each other if and only if at time t they are in the same g for some $g \in \mathcal{G}_t$. For an arbitrary $g \in \mathcal{G}_t$, we let $\mathfrak{A}_{t,g}^{u,\mathcal{G}}$ be the set of uninfected agents in g at time t . Accordingly, let $\Delta\mathfrak{A}_{t,g}^{\mathcal{G}}$ be the set of agents in $\mathfrak{A}_{t,g}^{u,\mathcal{G}}$ that become infected at $t + \Delta t$ under the slower coupled process. By Lemma 2.4.2, we have

$$\Pr \left[|\Delta\mathfrak{A}_{t,g}^{\mathcal{G}}| \geq \frac{\tau_0 \min\{\hat{\ell}_2/(\log^2 n), |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|\}}{4 \log^4 n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n. \quad (2.20)$$

Note that Lemma 2.4.2 requires that both the number of infected agents and the number of uninfected agents are at most $\hat{\ell}_2/(\log^2 n)$. By \mathcal{G}_t 's construction, there are

at least $\hat{\ell}_2/2$ infected agents in each subcube, and we may choose an arbitrary subset of them with size $\hat{\ell}_2/\log^2 n$ to form A^f in Lemma 2.4.2. We do not know the exact size of $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}|$ but in case $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}| > \hat{\ell}_2/\log^2 n$, we let A^f be an arbitrary subset of $\mathfrak{A}_{t,g}^{u,\mathcal{G}}$ with size $\hat{\ell}_2/\log^2 n$.

When $D_t = 1$, we have $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}| \leq 2\hat{\ell}_2 \log^2 n$. Therefore,

$$\min\{\hat{\ell}_2/\log^2 n, |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|\} \geq \min\left\{\frac{|\mathfrak{A}_{t,g}^{u,\mathcal{G}}|}{2\log^4 n}, |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|\right\} \geq |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|/(2\log^4 n).$$

Equation 2.20 can be rewritten as

$$\Pr \left[|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \geq \frac{\tau_0 |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|}{8\log^8 n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

This also provides a lower bound over the expectation of $|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}|$, i.e.,

$$\mathbb{E}[|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1] \geq \frac{\tau_0^2}{8} |\mathfrak{A}_{t,g}^{u,\mathcal{G}}| \log^{-14} n.$$

We also have

$$\begin{aligned} \mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1] &= \sum_g \mathbb{E}[|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1] \\ &\geq \sum_g \frac{\tau_0^2}{8} |\mathfrak{A}_{t,g}^{u,\mathcal{G}}| \log^{-14} n \\ &= \frac{\tau_0^2}{8} |\mathfrak{A}_t^{u,\mathcal{G}}| \log^{-14} n. \end{aligned} \tag{2.21}$$

Furthermore, by the way we design the coupled process, the random variables $|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}|$ are independent given $\mathcal{F}_t, D_t = 1$.

We consider two cases.

Case 1. There exists $g \in \mathcal{G}_t$ such that $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}| \geq |\mathfrak{A}_t^{u,\mathcal{G}}|/\log^{29} n$. In this case, we have

$$\Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \geq \frac{\tau_0 |\mathfrak{A}_t^{u,\mathcal{G}}|}{4\log^{38} n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \Pr \left[|\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \geq \frac{\tau_0 |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|}{4\log^9 n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

Case 2. For all $g \in \mathcal{G}_t$, $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}| < |\mathfrak{A}_t^{u,\mathcal{G}}|/\log^{29} n$. Observe, on the other hand, that $\sum_g |\mathfrak{A}_{t,g}^{u,\mathcal{G}}| = |\mathfrak{A}_t^{u,\mathcal{G}}|$. In this case, we have the summation $\sum_g |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|^2 < |\mathfrak{A}_t^{u,\mathcal{G}}|^2 \log^{-29} n$. (and it is maximized when every non-zero $|\mathfrak{A}_{t,g}^{u,\mathcal{G}}|$ is exactly $|\mathfrak{A}_t^{u,\mathcal{G}}| \log^{-29} n$). We therefore have

$$\sum_{g \in \mathcal{G}_t} |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|^2 \leq |\mathfrak{A}_t^{u,\mathcal{G}}| \log^{-29} n \sum_{g \in \mathcal{G}_t} |\mathfrak{A}_{t,g}^{u,\mathcal{G}}| = |\mathfrak{A}_t^{u,\mathcal{G}}|^2 \log^{-29} n. \quad (2.22)$$

Next, by Hoeffding's inequality (See, e.g., Theorem A.1.3), we have

$$\begin{aligned} & \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \leq \frac{\tau_0 |\mathfrak{A}_t^{u,\mathcal{G}}|}{4 \log^{38} n} \middle| \mathcal{F}_t, D_t = 1 \right] \\ & \leq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \leq \frac{\mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1]}{2} \middle| \mathcal{F}_t, D_t = 1 \right] \\ & = \Pr \left[\sum_g |\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \leq \frac{\mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1]}{2} \middle| \mathcal{F}_t, D_t = 1 \right] \\ & \leq 2 \exp \left(- \frac{2(\frac{1}{2} \mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \mid \mathcal{F}_t, D_t = 1])^2}{\sum_{g \in \mathcal{G}_t} |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|^2} \right) \\ & \quad \text{(apply Hoeffding's inequality; we have } |\Delta \mathfrak{A}_{t,g}^{\mathcal{G}}| \leq |\mathfrak{A}_{t,g}^{u,\mathcal{G}}|) \\ & \leq 2 \exp \left(- \frac{2 \frac{\tau_0^2}{16^2} |\mathfrak{A}_t^{u,\mathcal{G}}|^2 \log^{-28} n}{|\mathfrak{A}_t^{u,\mathcal{G}}|^2 \log^{-29} n} \right) \quad \text{(by Equation 2.21 and Equation 2.22)} \\ & = \exp(-\Theta(\log n)) \\ & \leq 1 - \tau_0 \log^{-6} n, \end{aligned}$$

for sufficiently large n .

Proving the inequality regarding $|\Delta \mathfrak{A}_t^{\mathcal{B}}|$ is similar. We provide it here for completeness, but less patient readers may simply skip this part. We first couple the diffusion problem with a slower process. First, all the infected agents in \mathcal{G}_t at time t cannot transmit virus. Second, agents in \mathcal{B}_t are able to transmit virus to each other if and only if at time t they are in the same b for some $b \in \mathcal{B}_t$. For an arbitrary $b \in \mathcal{B}_t$, we let $\mathfrak{A}_{t,b}^{f,\mathcal{B}}$ be the set of uninfected agents in b at time t . Accordingly, let $\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}$ be

the set of agents in $\mathfrak{A}_{t,b}^{f,\mathcal{B}}$ that becomes infected at $t + \Delta t$ under the coupled process. For technical reasons, we require the slower diffusion in the subcube b to halt when $|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}|$ becomes large, i.e., $|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| = |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|$. This added constraint $|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| = |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|$ allows us to apply Hoeffding's inequality in an easier manner.

When $D_t = 1$, $\mathfrak{A}_{t,b}^{f,\mathcal{B}} \leq 2\hat{\ell}_2 \log^2 n$ and $\min\{|\mathfrak{A}_{t,b}^{f,\mathcal{B}}|, \hat{\ell}_2(\log^2 n)\} \geq |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|/(2\log^4 n)$.

By Lemma 2.4.2, we have

$$\Pr \left[|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| \geq \frac{\tau_0 |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|}{8 \log^8 n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

Similar to the analysis of $\mathfrak{A}_{t,g}^{u,\mathcal{G}}$, this inequality holds because we can always restrict to a subset of agents if the number of infected/uninfected agents in the subcube is too large to meet the requirement in Lemma 2.4.2. We also have

$$\mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \mid \mathcal{F}_t, D_t = 1] = \sum_b \mathbb{E}[|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| \mid \mathcal{F}_t, D_t = 1] \geq \frac{\tau_0^2}{8} |\mathfrak{A}_t^{f,\mathcal{B}}| \log^{-14} n. \quad (2.23)$$

Furthermore, by the way we design the coupled process, the random variables $|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}|$ are independent given $\mathcal{F}_t, D_t = 1$.

We consider two cases.

Case 1. There exists an $b \in \mathcal{B}_t$ such that $|\mathfrak{A}_{t,b}^{f,\mathcal{B}}| \geq |\mathfrak{A}_t^{f,\mathcal{B}}|/\log^{29} n$. In this case, we have

$$\Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \geq \frac{\tau_0 |\mathfrak{A}_t^{f,\mathcal{B}}|}{4 \log^{38} n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \Pr \left[|\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| \geq \frac{\tau_0 |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|}{4 \log^9 n} \middle| \mathcal{F}_t, D_t = 1 \right] \geq \tau_0 \log^{-6} n.$$

Case 2. For all $b \in \mathcal{B}_t$, $|\mathfrak{A}_{t,b}^{f,\mathcal{B}}| < |\mathfrak{A}_t^{f,\mathcal{B}}|/\log^{29} n$. In this case, we have

$$\sum_{b \in \mathcal{B}_t} |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|^2 \leq |\mathfrak{A}_t^{f,\mathcal{B}}|^2 \log^{-29} n. \quad (2.24)$$

Next, by Hoeffding's inequality (again by Theorem A.1.3),

we have

$$\begin{aligned}
& \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \leq \frac{\tau_0 |\mathfrak{A}_t^{f,\mathcal{B}}|}{4 \log^{38} n} \middle| \mathcal{F}_t, D_t = 1 \right] \\
& \leq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \leq \frac{\mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \mid \mathcal{F}_t, D_T = 1]}{2} \middle| \mathcal{F}_t, D_t = 1 \right] \\
& = \Pr \left[\sum_b |\Delta \mathfrak{A}_{t,b}^{\mathcal{B}}| \leq \frac{\mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \mid \mathcal{F}_t, D_T = 1]}{2} \middle| \mathcal{F}_t, D_t = 1 \right] \\
& \leq 2 \exp \left(- \frac{2(\frac{1}{2} \mathbb{E}[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \mid \mathcal{F}_t, D_t = 1])^2}{\sum_{b \in \mathcal{B}_t} |\mathfrak{A}_{t,b}^{f,\mathcal{B}}|^2} \right) \\
& \quad \text{(apply Hoeffding's inequality; we have } |\Delta \mathfrak{A}_{t,g}^{\mathcal{B}}| \leq |\mathfrak{A}_{t,g}^{f,\mathcal{B}}|. \text{)} \\
& \leq 2 \exp \left(- \frac{2 \frac{\tau_0^4}{16^2} |\mathfrak{A}_t^{f,\mathcal{B}}|^2 \log^{-28} n}{|\mathfrak{A}_t^{f,\mathcal{B}}|^2 \log^{-29} n} \right) \quad \text{(by Equation 2.23 and Equation 2.24)} \\
& \leq 1 - \tau_0 \log^{-6} n.
\end{aligned}$$

□

2.4.1 Leveraging local analysis

We now move to the global diffusion upper bound. As discussed in the beginning of this section, the balance between the distributions of each type of subcube and the distributions of actual agents plays a crucial role in our analysis. Fix an arbitrary time t , we classify the joint configurations of the agents into four types:

- type 1 (namely $\mathcal{P}_{1,t}$): when $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^f|$.
- type 2 (namely $\mathcal{P}_{2,t}$): when $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^f|$.
- type 3 (namely $\mathcal{P}_{3,t}$): when $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^u|$.
- type 4 (namely $\mathcal{P}_{4,t}$): when $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^u|$.

Recall that \mathcal{F}_t refers to the information on the global configurations up to time t . We shall abuse notation slightly and say $\mathcal{F}_t \in \mathcal{P}_{i,t}$ if the configuration of the agents at time t belongs to the i th type described above. Notice that \mathcal{F}_t belongs to exactly one of the sets $\mathcal{P}_{1,t}, \mathcal{P}_{2,t}, \mathcal{P}_{3,t}, \mathcal{P}_{4,t}$. In brief, scenarios $\mathcal{P}_{1,t}$ and $\mathcal{P}_{2,t}$ have a majority of uninfected subcubes, while $\mathcal{P}_{3,t}$ and $\mathcal{P}_{4,t}$ have a majority of infected subcubes. From another perspective, $\mathcal{P}_{1,t}$ and $\mathcal{P}_{3,t}$ refer to situations when the dominant types (with respect to the status of infection) are dense in their subcube types (infected/uninfected subcubes), while $\mathcal{P}_{2,t}$ and $\mathcal{P}_{4,t}$ refer to the more uniform scenarios. The next lemma states that when $\mathcal{F}_t \in \mathcal{P}_{1,t} \cup \mathcal{P}_{2,t}$, the total number of infected agents $|\mathfrak{A}_t^f|$ grows in proportion to a monotone function of $|\mathfrak{A}_t^f|$ within Δt steps. On the other hand, when $\mathcal{F}_t \in \mathcal{P}_{3,t} \cup \mathcal{P}_{4,t}$, the total number of uninfected agents $|\mathfrak{A}_t^u|$ is reduced in proportion to a monotone function of $|\mathfrak{A}_t^u|$ within Δt steps.

Lemma 2.4.9. *Fix an arbitrary t , define the following events,*

$$\begin{aligned} e_1(t) &= \left\{ |\Delta \mathfrak{A}_t| \geq 0.09\tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\ell_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \right\} & e_2(t) &= \left\{ |\Delta \mathfrak{A}_t| \geq \frac{\tau_0^2}{8 \log^{38} n} |\mathfrak{A}_t^f| \right\} \\ e_3(t) &= \left\{ |\Delta \mathfrak{A}_t| \geq 0.015\tau_0 \left(\frac{|\mathfrak{A}_t^u|}{4\ell_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \right\} & e_4(t) &= \left\{ |\Delta \mathfrak{A}_t| \geq \frac{\tau_0^2}{8 \log^{38} n} |\mathfrak{A}_t^u| \right\}. \end{aligned}$$

We have

$$\Pr[e_i \mid \mathcal{F}_t \in \mathcal{P}_{i,t}, D_t = 1] \geq \tau_0 \log^{-6} n$$

for $i = 1, 2, 3, 4$.

Intuitively, e_1 and e_2 connect the number of newly infected agents to the original number of infected agents. When e_1 or e_2 are triggered sufficiently many times, the number of infected agents doubles. Meanwhile, e_3 and e_4 connect the number of newly infected agents to the original number of uninfected agents. When e_3 or e_4 are triggered sufficiently many times, the number of uninfected agents halves.

The key to proving Lemma 2.4.9, which will ultimately lead to a bound on the global growth rate of doubling/halving the total number of infected/uninfected agents as depicted in the next proposition, is a geometric relation between the boundary of \mathcal{G}_t , i.e. $\partial\mathcal{G}_t$, and \mathcal{G}_t itself. More specifically, an isoperimetric bound on \mathcal{G}_t guarantees that no matter how packed together these good subcubes are, there are still an order $|\mathcal{G}_t|^{2/3}$ of them exposed to the bad subcubes, hence the global infection rate cannot be too slow.

Let us start with the isoperimetric bound we need.

Theorem 2.4.10. *Let \mathcal{G} be an arbitrary subset of \mathfrak{C}^d . There exists a pair of constants $\alpha(d) > 1/2$ and $\beta(d) > 0$, such that:*

$$\text{if } |\mathcal{G}| \leq \alpha(d) \cdot |\mathbf{Z}^d| = \alpha(d) \cdot b^d, \quad \text{then } |\partial\mathcal{G}| \geq \beta(d) |\mathcal{G}|^{(d-1)/d}$$

Specifically, $\beta(3) \geq 0.36$.

The isoperimetric problem over \mathfrak{C}^d was studied in [16], in which the optimal structure of \mathcal{G} that minimizes $|\partial\mathcal{G}|$ is presented. Here, we provide another asymptotically optimal proof based on a recursive argument. This proof could be of independent interest.

To begin, let us prove the special case $d = 2$. The analysis for this case demonstrates important ideas that are needed for showing the case for general d .

Lemma 2.4.11. *Let \mathcal{G} be an arbitrary subset of \mathfrak{C}^2 . If $|\mathcal{G}| \leq \frac{2}{3}b^2$, we have*

$$|\partial\mathcal{G}| \geq \frac{2}{5} |\mathcal{G}|^{1/2}.$$

Proof. Let $V = |\mathcal{G}|$ and $X(i)$ be the collection of lattice points in \mathfrak{E}^2 whose x coordinates are i . Also we refer $V(i) := X(i) \cap \mathcal{G}$ as the i th *stripe* of \mathcal{G} . Define

$$i^* = \arg \max_i |V(i)| \quad \text{and} \quad i_* = \arg \min_i |V(i)|.$$

We next analyze two possible cases regarding the size of $V(i^*)$.

Case 1. $0 < |V(i^*)| < \sqrt{\frac{3V}{2}}$. Since $\sqrt{\frac{3V}{2}} \leq b$, for each i such that $V(i) \neq \emptyset$, we have

$$0 < |V(i)| \leq |V(i^*)| < b.$$

On the other hand, when $0 < |V(i)| < b$, there is at least one element of $X(i)$ that is also in $\partial\mathcal{G}$. Since the cardinality of \mathcal{G} is V , the number of non-empty stripes in \mathcal{G} is at least $\frac{V}{|V(i^*)|}$. Hence we have

$$|\partial\mathcal{G}| \geq \frac{V}{|V(i^*)|} \geq \sqrt{\frac{2V}{3}} > \frac{2}{5}\sqrt{V}$$

Case 2. $|V(i^*)| \geq \sqrt{\frac{3V}{2}}$. By an averaging argument, $|V(i_*)| \leq V/b$. Using the fact that $V \leq \frac{2}{3}b^2$, we have $|V(i_*)| \leq \sqrt{\frac{2}{3}}V$.

Next we show that $|\partial\mathcal{G}| \geq |V(i^*)| - |V(i_*)|$. Consider an arbitrary j such that $(i^*, j) \in V(i^*)$ and $(i_*, j) \notin V(i_*)$. Since $(i^*, j) \in \mathcal{G}$ and $(i_*, j) \notin \mathcal{G}$, there exists a lattice point on the “line segment” $\{(i, j) : i \in \{i^*, \dots, i_*\}\}$ that is in $\partial\mathcal{G}$.

Finally, we have

$$|\partial\mathcal{G}| \geq |V(i^*)| - |V(i_*)| \geq \left(\sqrt{\frac{3}{2}} - \sqrt{\frac{2}{3}} \right) \sqrt{V} \geq \frac{2}{5}\sqrt{V}.$$

□

We use induction to prove Theorem 2.4.10. Our idea of proving general d is similar to the case $d = 2$. First, we let $X(i)$ be the collection of lattice points

in \mathfrak{C}^d whose first coordinates are i and $V(i) = X(i) \cap \mathcal{G}$. Next, we also define $i^* = \arg \max_i |V(i)|$ and $i_* = \arg \min_i |V(i)|$. Then, we mimic the analysis for the case $d = 2$ and discuss two possible cases: when $|V(i^*)|$ is small and when $|V(i^*)|$ is large. When $|V(i^*)|$ is small, we need to invoke results on lower dimension cases. When $|V(i^*)|$ is large, we shall show that $|V(i^*)| - |V(i_*)|$ is a lower bound on the size of ∂G , which is sufficient for proving the theorem.

Let us proceed with the following lemma, which is the main vehicle for analyzing the case $|V(i^*)|$ is large.

Lemma 2.4.12. *Let \mathcal{G} be an arbitrary subset of \mathfrak{C}^d . We have*

$$|\partial \mathcal{G}| \geq |V(i^*)| - |V(i_*)|.$$

Proof. First, define the set Δ as

$$\Delta = \left\{ (i_2, i_3, \dots, i_d) \in \mathfrak{C}^{d-1} \mid ((i^*, i_2, i_3, \dots, i_d) \in V(i^*)) \wedge ((i_*, i_2, i_3, \dots, i_d) \notin V(i_*)) \right\}$$

Notice that by the definitions of $V(i^*)$ and $V(i_*)$, we have $|\Delta| \geq |V(i^*)| - |V(i_*)|$. Next, we show that for any $(i_2, \dots, i_d) \in \Delta$, there exists an i_1 such that $(i_1, \dots, i_d) \in \partial \mathcal{G}$, which immediately implies the lemma.

Fix a $(d-1)$ -tuple $(i_2, \dots, i_d) \in \Delta$. Observe that $(i^*, i_2, \dots, i_d) \in V(i^*) \subseteq \mathcal{G}$ and $(i_*, i_2, \dots, i_d) \notin V(i_*)$ and thus $(i_*, i_2, \dots, i_d) \notin \mathcal{G}$. Let us walk from the point (i^*, i_2, \dots, i_d) to the point (i_*, i_2, \dots, i_d) . Because we start with an interior point of \mathcal{G} and end at a point outside \mathcal{G} , we leave the polytope \mathcal{G} at least once. Hence, there exists an i_1 such that $(i_1, \dots, i_d) \in \partial G$. \square

Now we are ready to prove Theorem 2.4.10.

Proof of Theorem 2.4.10. We prove by induction on d . Specifically, we show that for any d and any $\mathcal{G}(d) \subseteq \mathfrak{C}^d$, there exists a pair of constants (that depends only on d) $\alpha(d) \geq 1/2$ and $\beta(d) > 0$ such that

$$\text{if } |\mathcal{G}| \leq \alpha(d)|\mathfrak{C}^d|, \quad \text{then } |\partial\mathcal{G}(d)| \geq \beta(d)|\mathcal{G}|^{d-1}.$$

The base case was considered in Lemma 2.4.11. Now let us assume the theorem holds up to the d -dimensional space. We now prove the $d + 1$ dimensional case.

Our $\alpha(d + 1)$ and $\beta(d + 1)$ are set up in the following way:

$$\begin{aligned} \alpha(d + 1) &= \alpha(d)/2 + 1/4 \\ \beta(d + 1) &= \min \left\{ \frac{\alpha(d)}{(\alpha(d+1))^{\frac{d}{d+1}}} - (\alpha(d + 1))^{\frac{1}{d+1}}, \frac{\beta(d)(\alpha(d+1))^{\frac{1}{d+1}}}{(\alpha(d))^{\frac{1}{d}}} \right\} \end{aligned} \quad (2.25)$$

Let $T = \left(\frac{\alpha(d)}{(\alpha(d+1))^{\frac{d}{d+1}}} \right) V^{\frac{d}{d+1}}$ and consider the following two cases.

Case 1. $|V(i^*)| < T$. Our $\alpha(d + 1)$ is set up in a way that when $V \leq \alpha(d + 1)b^{d+1}$, $T < \alpha(d)b^d$. Next, we invoke the result for d dimensional case on each $V(i)$, $i \in [b]$. Notice that a lattice on the exterior surface of $V(i)$ in the space \mathfrak{C}^d is also on the exterior surface of \mathcal{G} . Let us call the size of the exterior surface of $V(i)$ as $|\partial V(i)|$. By induction hypothesis, we have $|\partial V(i)| \geq \beta(d)|V(i)|^{\frac{d-1}{d}}$. Note also $\sum_{i \leq b} |V(i)| = V$.

Next, define $f(x) = x^{\frac{d-1}{d}}$, which is a concave function. We have

$$\begin{aligned}
|\partial\mathcal{G}| &\geq \sum_{i \leq b} |\partial V(i)| \\
&\geq \sum_{i \leq b} \beta(d) f(|V(i)|) \quad (\text{induction hypothesis}) \\
&\geq \sum_{i \leq b} \frac{\beta(d) |V(i)|}{T} f(T) \quad (|V(i^*)| < T \text{ and using the concave property of } f(\cdot)) \\
&= \frac{\beta(d) V}{T} f(T) \\
&= \frac{\beta(d) (\alpha(d+1))^{\frac{1}{d+1}}}{(\alpha(d))^{\frac{1}{d}}} V^{\frac{d}{d+1}} \quad (\text{using the definition of } T) \\
&\geq \beta(d+1) V^{\frac{d}{d+1}} \quad (\text{by the construction of } \beta(d))
\end{aligned}$$

Case 2. When $|V(i^*)| \geq T$. By Lemma 2.4.12, $|\partial\mathcal{G}| \geq |V(i^*)| - |V(i_*)|$. Also by an averaging argument we have $|V(i_*)| \leq V/b$. The theorem then follows. \square

Proof of Lemma 2.4.9. Part 1. $\mathcal{P}_{1,t}$, $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^f|$. Since $D_t = 1$, the number of agents in each subcube is at most $2\hat{\ell}_2 \log^2 n$. Therefore, $|\mathcal{G}_t| \geq |\mathfrak{A}_t^{f,\mathcal{G}}|/(2\hat{\ell}_2 \log^2 n)$. To apply Corollary 2.4.6, we need to derive a relationship between the size of \mathcal{G}_t and the size of $\partial\mathcal{G}_t$. This is an isoperimetric problem. By Theorem 2.4.10,

$$|\partial\mathcal{G}_t| \geq 0.36 |\mathcal{G}_t|^{2/3} \geq 0.36 \left(\frac{|\mathfrak{A}_t^{f,\mathcal{G}}|}{2\hat{\ell}_2 \log^2 n} \right)^{2/3}.$$

We have

$$\begin{aligned}
& \Pr \left[|\Delta \mathfrak{A}_t| \geq 0.09\tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{1,t}, D_t = 1 \right] \\
& \geq \Pr \left[|\widetilde{\Delta \mathfrak{A}_t^{\mathcal{G}}}| \geq 0.09\tau_0 \left(\frac{|\mathfrak{A}_t^{f,\mathcal{G}}|}{2\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{1,t}, D_t = 1 \right] \\
& \geq \Pr \left[|\widetilde{\Delta \mathfrak{A}_t^{\mathcal{G}}}| \geq |\partial \mathcal{G}_t| \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{1,t}, D_t = 1 \right] \\
& \geq \tau_0 \log^{-6} n \quad (\text{by Corollary 2.4.6})
\end{aligned}$$

Part 2. $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^f|$. Notice that $|\mathfrak{A}_t^{f,\mathcal{B}}| \geq |\mathfrak{A}_t^f|/2$ and $|\Delta \mathfrak{A}_t^{\mathcal{B}}| \leq |\Delta \mathfrak{A}_t|$. We have

$$\begin{aligned}
& \Pr \left[|\Delta \mathfrak{A}_t| \geq \frac{\tau_0^2}{8 \log^{38} n} |\mathfrak{A}_t^f| \middle| \mathcal{F}_t \in \mathcal{P}_{2,t}, D_t = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \geq \frac{\tau_0^2}{4 \log^{38} n} |\mathfrak{A}_t^{f,\mathcal{B}}| \middle| \mathcal{F}_t \in \mathcal{P}_{2,t}, D_t = 1 \right] \\
& \geq \tau_0 \log^{-6} n \quad (\text{by Corollary 2.4.7})
\end{aligned}$$

Part 3. $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^u|$. This is similar to part 1. We have

$$|\partial \mathcal{G}_t| = |\partial \mathcal{B}_t| \geq |\partial \mathcal{B}_t|/6 \geq 0.06|\mathcal{B}_t|^{2/3} \geq 0.06 \left(\frac{|\mathfrak{A}_t^{u,\mathcal{B}}|}{2\hat{\ell}_2 \log^2 n} \right)^{2/3}.$$

The second inequality holds because the exterior surface $\partial \mathcal{B}_t$ is in the neighborhood of $\partial \mathcal{B}_t$ and $|N(\partial \mathcal{B}_t)| \leq 6|\partial \mathcal{B}_t|$. Notice that $|\mathfrak{A}_t^{u,\mathcal{B}}| \geq |\mathfrak{A}_t^u|/2$. We have

$$\begin{aligned}
& \Pr \left[|\Delta \mathfrak{A}_t| \geq 0.015\tau_0 \left(\frac{|\mathfrak{A}_t^u|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{3,t}, D_t = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \geq 0.015\tau_0 \left(\frac{|\mathfrak{A}_t^{u,\mathcal{B}}|}{2\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{3,t}, D_t = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{B}}| \geq |\partial \mathcal{G}_t| \frac{\tau_0 \hat{\ell}_2}{4 \log^{13} n} \middle| \mathcal{F}_t \in \mathcal{P}_{3,t}, D_t = 1 \right] \\
& \geq \tau_0 \log^{-6} n \quad (\text{by Corollary 2.4.6})
\end{aligned}$$

Part 4. $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^u|$. This is similar to part 2.

Notice that $|\mathfrak{A}_t^u| \leq 2|\mathfrak{A}_t^{u,\mathcal{G}}|$ and $|\Delta \mathfrak{A}_t| \geq |\Delta \mathfrak{A}_t^{\mathcal{G}}|$. We have

$$\begin{aligned} & \Pr \left[|\Delta \mathfrak{A}_t| \geq \frac{\tau_0^2}{8 \log^{38} n} |\mathfrak{A}_t^u| \mid \mathcal{F}_t \in \mathcal{P}_{4,t}, D_t = 1 \right] \\ & \geq \Pr \left[|\Delta \mathfrak{A}_t^{\mathcal{G}}| \geq \frac{\tau_0^2}{4 \log^{38} n} |\mathfrak{A}_t^{u,\mathcal{G}}| \mid \mathcal{F}_t \in \mathcal{P}_{4,t}, D_t = 1 \right] \\ & \geq \tau_0 \log^{-6} n. \quad (\text{by Corollary 2.4.7}) \end{aligned}$$

□

Our major proposition presented next essentially pins down the number of times these events need to be triggered to double the number of infected agents or halve the number of uninfected ones.

Proposition 2.4.13. *Consider the information diffusion problem over \mathcal{V}^3 with m agents. For any fixed $t \leq n^{2.5} - 4\sqrt{\frac{m}{n}} \log^{45} n \Delta t$, define the following events*

$$\chi_1(t) \equiv \left(|\mathfrak{A}_{t+4\sqrt{\frac{m}{n}} \log^{45} n \Delta t}^f| \geq 2|\mathfrak{A}_t^f| \right) \quad \text{and} \quad \chi_2(t) \equiv \left(|\mathfrak{A}_{t+4\sqrt{\frac{m}{n}} \log^{45} n \Delta t}^u| \leq \frac{1}{2}|\mathfrak{A}_t^u| \right).$$

We have

$$\Pr[\chi_1(t) \vee \chi_2(t)] \geq 1 - \exp(-\log^2 n).$$

Note that this bound suggests that for each time increment $4\sqrt{\frac{m}{n}} \log^{45} n \Delta t$, either the number of infected agents doubles or the number of uninfected agents is reduced by half with high probability. Therefore, within time at most $2 \log n \cdot (4\sqrt{\frac{m}{n}} \log^{45} n \Delta t) = 128n\hat{\ell}_2 \log^{47} n$ all the agents get infected with probability at least $1 - 2 \log n \exp(-\log^2 n)$. This proves Theorem 2.4.1.

To summarize our approach, Corollaries 2.4.6 and 2.4.7 first translate the local infection rate of Lemma 2.4.2 into a rate based on the subcube types (i.e. good and

bad subcubes). Then Lemma 2.4.9 further aggregates the growth rate to depend only on the infected and uninfected agents, by looking at the geometrical arrangement of the subcubes. Nevertheless, the bound from Lemma 2.4.9 is still too crude, but by making a long enough sequence of trials i.e. $4\sqrt{\frac{m}{n}} \log^{45} n$ times, at least one of the four scenarios defined in Lemma 2.4.9 occurs for a significant number of times, despite the $\Omega(\log^{-6} n)$ probability of occurrence for each individual step for any of the four scenarios. This leads to the probabilistic bound for $\chi_1(t) \vee \chi_2(t)$.

Proof of Proposition 2.4.13. First, notice that Lemma 2.4.9 states that regardless of the diffusion process' history, one of e_1, e_2, e_3 , and e_4 is guaranteed to occur with $\tilde{\Omega}(1)$ probability. On the other hand, it is not difficult to see that when *any* of the events e_1, e_2, e_3, e_4 occurs $\tilde{\Omega}(n/\hat{\ell}_2)$ times, then either $|\mathfrak{A}_t^f|$ doubles or $|\mathfrak{A}_t^u|$ reduces by a half. Although we do not know exactly which event happens at a specific time t , we argue that so long as we wait long enough, the collection of events $\{e_1, \dots, e_4\}$ occurs $4\tilde{\Omega}(n/\hat{\ell}_2)$ times and by Pigeonhole principle, at least one of e_1, e_2, \dots, e_4 will be triggered $\tilde{\Omega}(n/\hat{\ell}_2)$ times, concluding the proposition. The above argument can be made rigorous via a Chernoff bound on the *total* number of occurrence of all four events.

We now implement the idea. Let us define $\varsigma = 4\sqrt{\frac{m}{n}} \log^{45} n$. Let $t_i = t + (i-1)\Delta t$ for $i \in [\varsigma]$. Note that t_i depends on both i and t , but we suppress the dependence on t for succinctness. This also applies to all other defined quantities in this proof. Recall from Lemma 2.3.13 that \mathcal{F}_{t_i} encodes all the available information up to time

t_i . For each $i \in [\varsigma]$, define the following pairs of indicator functions

$$I_{1,2}(t_i) = \begin{cases} 1 & \text{if } \mathcal{F}_{t_i} \in \mathcal{P}_{t_i,1} \cup \mathcal{P}_{t_i,2} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad I_{3,4}(t_i) = \begin{cases} 1 & \text{if } \mathcal{F}_{t_i} \in \mathcal{P}_{t_i,3} \cup \mathcal{P}_{t_i,4} \\ 0 & \text{otherwise} \end{cases}$$

Notice that for arbitrary t_i , $I_{1,2}(t_i) + I_{3,4}(t_i) = 1$. Next define

$$\varphi_i = I_{1,2}(t_i) \cdot \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^f|} + I_{3,4}(t_i) \cdot \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^u|},$$

We first show a lower bound for $\sum_{i \leq \varsigma} \varphi_i$. Our strategy is to invoke Lemma 2.4.9 and apply a Chernoff bound. Special care needs to be taken when $D = 0$.

Let

$$r = \min \left\{ 0.015\tau_0 \left(\frac{\hat{\ell}_2}{4m \log^2 n} \right)^{\frac{1}{3}} \frac{1}{\log^{16} n}, \frac{\tau_0^2}{8 \log^{38} n} \right\}.$$

By Lemma 2.4.9, we can see regardless of whether \mathcal{F}_{t_i} belongs to $\mathcal{P}_{1,t_i}, \dots$, or \mathcal{P}_{4,t_i} ,

$$\Pr[\varphi_i \geq r \mid \mathcal{F}_{t_i}, D_{t_i} = 1] \geq \tau_0 \log^{-6} n. \quad (2.26)$$

where τ_0 is the constant specified in Lemma 2.4.9. Here, we verify the case $\mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}$ for Equation 2.26. See (2.27). The computation for the other three cases can be carried out similarly.

Next, let us define a family of indicator random variables $\{I(i) : i \leq \varsigma\}$ so that $I(i)$ is \mathcal{F}_{t_i} -measurable and

$$I(i) = \begin{cases} 1 & \text{if } \varphi_i \geq r \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $\sum_{i \leq \varsigma} \varphi_i \geq r \left(\sum_{i \leq \varsigma} I(i) \right)$. By Equation 2.26, we have

$$\Pr[I(i) = 1 \mid \mathcal{F}_{t_i}, D_{t_i} = 1] \geq \tau_0 \log^{-6} n.$$

$$\begin{aligned}
& \Pr[\varphi_i \geq r \mid \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1] \\
& \geq \Pr \left[\frac{|\Delta \mathfrak{A}_t|}{|\mathfrak{A}_t^f|} \geq \frac{0.015}{\log^{16} n} \tau_0 \left(\frac{\hat{\ell}_2}{4m \log^2 n} \right)^{\frac{1}{3}} \middle| \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t| \geq \frac{0.09}{\log^{16} n} \tau_0 |\mathfrak{A}_t^f| \left(\frac{\hat{\ell}_2}{4m \log^2 n} \right)^{\frac{1}{3}} \middle| \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t| \geq \frac{0.09}{\log^{16} n} \tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \underbrace{|\mathfrak{A}_t^f|}_{\leq m}^{1/3} (4\hat{\ell}_2 \log^2 n)^{2/3} \left(\frac{\hat{\ell}_2}{4m \log^2 n} \right)^{\frac{1}{3}} \middle| \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1 \right] \quad (2.27) \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t| \geq \frac{0.09}{\log^3 n} \tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} m^{1/3} (4 \log^2 n)^{2/3} \left(\frac{1}{4m \log^2 n} \right)^{\frac{1}{3}} \middle| \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1 \right] \\
& \geq \Pr \left[|\Delta \mathfrak{A}_t| \geq 0.09 \tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \middle| \mathcal{F}_{t_i} \in \mathcal{P}_{1,t_i}, D_{t_i} = 1 \right] \\
& \geq \tau_0 \log^{-6} n \text{ (Lemma 2.4.9)}
\end{aligned}$$

Next, let us introduce another family of r.v. $\{I'(i) : i \leq \varsigma\}$ to incorporate the good density variable as follows:

$$I'(i) = \begin{cases} I(i) & \text{if } D_{t_i} = 1 \\ 1 & \text{otherwise} \end{cases}$$

Since $I'(i) \geq I(i)$, we also have

$$\mathbb{E}[I'(i) \mid \mathcal{F}_{t_i}, D_{t_i} = 1] \geq \tau_0 \log^{-6} n.$$

On the other hand, by construction

$$\mathbb{E}[I'(i) \mid \mathcal{F}_{t_i}, D_{t_i} = 0] = 1 \geq \log^{-7} n$$

This concludes that

$$\mathbb{E}[I'(i) \mid \mathcal{F}_{t_i}] \geq \log^{-7} n$$

which implies

$$\mathbb{E}\left[\sum_{i \leq \varsigma} I'(i)\right] \geq \varsigma \log^{-7} n.$$

We construct a sequence of random variables $\{\xi_i\}$ such that $\xi_0 = 0$ and $\xi_{i+1} = \xi_i + (I'(i+1) - \mathbb{E}[I'(i+1) \mid \mathcal{F}_{t_i}])$. We can verify that ξ_i is a martingale with respect to $\{\mathcal{F}_{t_i}\}$ and $|\xi_i - \xi_{i-1}| \leq 2$ for all i . By Azuma-Hoeffding's inequality (see Theorem A.1.3),

$$\begin{aligned} & \Pr \left[|\xi_\varsigma| \geq \frac{1}{2} \mathbb{E}\left[\sum_{i \leq \varsigma} I'(i)\right] \right] \\ & \leq 2 \exp \left(-\frac{\frac{1}{4} \mathbb{E}^2\left[\sum_{i \leq \varsigma} I'(i)\right]}{2 \sum_{i \leq \varsigma} 4} \right) \\ & \leq 2 \exp \left(-\frac{1}{32} \varsigma \log^{-14} n \right) \leq \exp(-\log^{30} n). \end{aligned}$$

This implies

$$\Pr\left[\sum_{i \leq \varsigma} I'(i) \leq \frac{1}{4}\varsigma\tau_0 \log^{-6} n\right] \leq \exp(-\log^{30} n).$$

Next, notice that

$$\Pr\left[\sum_{i \leq \varsigma} I'(i) \neq \sum_{i \leq \varsigma} I(i)\right] \leq \Pr[D_{t+(\varsigma-1)\Delta t} = 0] \leq \Pr[D = 0] \leq \exp(-\frac{1}{15} \log^2 n)$$

where the last inequality follows from Lemma 2.4.5.

We conclude that

$$\Pr\left[\sum_{i \leq \varsigma} I(i) \leq \frac{1}{4}\varsigma\tau_0 \log^{-6} n\right] \leq \exp(-\log^{30} n) + \exp(-\frac{1}{15} \log^2 n) \leq 2 \exp(-\frac{1}{15} \log^2 n).$$

Finally, we show when $(\sum_{i \leq \varsigma} I(i) > \frac{1}{4}\varsigma\tau_0 \log^{-6} n)$ occurs, either $\chi_1(t)$ or $\chi_2(t)$ is true. First, we have a lower bound $\sum_{i \leq \varsigma} \varphi_i \geq r \cdot \frac{\tau_0 \varsigma}{4 \log^6 n} \geq 2$. Next, we show this lower bound results in a minimum guarantee on $|\Delta \mathfrak{A}_t|$. Specifically, we have

$$\left(\sum_{i \leq \varsigma} I_{1,2}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^f|}\right) + \left(\sum_{i \leq \varsigma} I_{3,4}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^u|}\right) \geq 2,$$

which implies either

$$\sum_{i \leq \varsigma} I_{1,2}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^f|} \geq 1 \text{ (Case 1)}$$

or

$$\sum_{i \leq \varsigma} I_{3,4}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^u|} \geq 1 \text{ (Case 2)}$$

Case 1. Observe that

$$|\mathfrak{A}_{t+4\sqrt{\frac{m}{n}} \log^{45} \Delta t}^f - \mathfrak{A}_t^f| = \sum_{i \leq \varsigma} |\Delta \mathfrak{A}_{t_i}|$$

because $\Delta \mathfrak{A}_{t_i}^f$ are all disjoint for different i . We have

$$\begin{aligned}
& |\mathfrak{A}_{t+4\sqrt{\frac{m}{n}}\log^{45}\Delta t}^f - \mathfrak{A}_t^f| \\
& \geq \sum_{i \leq \varsigma} |\Delta \mathfrak{A}_{t_i}| \\
& \geq \sum_{i \leq \varsigma} I_{1,2}(t_i) |\Delta \mathfrak{A}_{t_i}| \\
& = |\mathfrak{A}_t^f| \sum_{i \leq \varsigma} I_{1,2}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_t^f|} \\
& \geq |\mathfrak{A}_t^f| \left(\sum_{i \leq \varsigma} I_{1,2}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^f|} \right) \quad (|\mathfrak{A}_t^f| \text{ is non decreasing w.r.t. } t) \\
& \geq |\mathfrak{A}_t^f|.
\end{aligned}$$

In this case, the event $\chi_1(t)$ occurs.

Case 2. When $\mathfrak{A}_{t_\varsigma}^u = \emptyset$, nothing needs to be proved. Let us focus on the situation where $\mathfrak{A}_{t_\varsigma}^u \neq \emptyset$

$$\begin{aligned}
& |\mathfrak{A}_{t+4\sqrt{\frac{m}{n}}\log^{45}\Delta t}^u - \mathfrak{A}_t^u| \\
& \geq \sum_{i \leq \varsigma} |\Delta \mathfrak{A}_{t_i}| \\
& \geq \sum_{i \leq \varsigma} I_{3,4}(t_i) |\Delta \mathfrak{A}_{t_i}| \\
& = |\mathfrak{A}_{t_\varsigma}^u| \sum_{i \leq \varsigma} I_{3,4}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_\varsigma}^u|} \\
& \geq |\mathfrak{A}_{t_\varsigma}^u| \left(\sum_{i \leq \varsigma} I_{3,4}(t_i) \frac{|\Delta \mathfrak{A}_{t_i}|}{|\mathfrak{A}_{t_i}^u|} \right) \quad (|\mathfrak{A}_t^u| \text{ is non increasing w.r.t. } t) \\
& \geq |\mathfrak{A}_{t_\varsigma}^u|.
\end{aligned}$$

Therefore, the event $\chi_2(t)$ occurs in this case.

□

2.5 The case when the number of agents is sparse

This section focuses on the case where $m = o(n)$:

Proposition 2.5.1. *Let a_1, a_2, \dots, a_m be placed uniformly at random on \mathcal{V}^3 , where $m < n \log^{-2} n$. Let a_1 be the agent that holds a virus at $t = 0$, and T be the diffusion time. We have for any constant $c > 0$,*

$$\Pr[T < \frac{n^3}{m} \log^{-c} n] \leq \log^{-c} n$$

and

$$\Pr[T > \frac{2n^3}{m} \log^{15} n] \leq \exp(-(\log^2 n)/2).$$

Note that our analysis in Chapter 2.3 and Chapter 2.4 cannot be applied directly to prove Proposition 2.5.1 because we required the side of each subcube to be of length $\ell_2 = \sqrt{\frac{n^3}{m}}$, which is larger than $(2n + 1)$ when $m = o(n)$. The diffusion time for this case turns out to depend on m and n in a way different from the case where $n \log^2 n < m < n^3$. Nevertheless, some of the arguments can still be borrowed from Lemma 2.4.2, together with the use of mixing time of a random walk in \mathcal{V}^3 . Because of the similarity of our analysis with previous sections, we only sketch our proof and highlight the new main technicalities.

We first show the lower bound of the diffusion time:

Lemma 2.5.2. *Let a_1, a_2, \dots, a_m be placed uniformly at random on \mathcal{V}^3 , where $m < 2n + 1$. Let a_1 be the agent that holds a virus at $t = 0$. Let T be the diffusion time. We have, for any constant $c > 0$,*

$$\Pr[T < \frac{n^3}{m} \log^{-c} n] \leq \log^{-c} n$$

Proof. Let these m random walks be S^1, S^2, \dots, S^m . Since each random walk is already at stationary distribution at $t = 0$, they are all distributed uniformly at any specific time. Therefore, for any fixed t and fixed $j > 1$, $\Pr[\|S_t^1 - S_t^j\|_1 \leq 1] \leq 7/(2n+1)^3$. By a union bound,

$$\Pr[\exists t \leq \frac{n^3}{m} \log^{-c} n, i > 1 : \|S_t^1 - S_t^i\|_1 \leq 1] \leq \frac{n^3}{m \log^c n} \cdot m \cdot 7(2n+1)^{-3} < \log^{-c} n.$$

Therefore, with probability at least $1 - \log^{-c} n$, S^1 will not meet any other agent before $t = \frac{n^3}{m} \log^{-c} n$, which also implies that the diffusion process has not been completed. \square

Next we move to the upper bound:

Lemma 2.5.3. *Let a_1, a_2, \dots, a_m be placed uniformly at random on \mathcal{V}^3 , where $m < \frac{n}{\log^2 n}$. Let a_1 be the agent that holds a virus at $t = 0$. Let T be the diffusion time. We have*

$$\Pr[T > \frac{2n^3}{m} \log^{15} n] \leq \exp(-(\log^2 n)/2).$$

The following is a key lemma for the upper bound analysis. The lemma reuses arguments that appeared in Lemma 2.4.2. However, as the agents are sparser in this case, new diffusion rules for the coupling process and the corresponding probabilistic bounds are needed.

Lemma 2.5.4. *Consider the diffusion process in which $m < \frac{n}{\log^2 n}$. Fix a time t , and let A^f and A^u be the set of infected and uninfected agents at time t with $|A^f| = m_1$ and $|A^u| = m_2$. Let c be a sufficiently large constant and $\Delta t = cn^3(\log n)/m$. Let $M(t)$ be the number of newly infected agents from time t to $t + \Delta t$. Assume the agents*

are arbitrarily (in an adversarial manner) distributed at time t . We have

$$\Pr \left[M(t) \geq \frac{\min\{m_1, m_2\}}{\log^5 n} \right] \geq \frac{1}{2} \log^{-5} n.$$

Proof. Similar to the proof of Lemma 2.4.2, we first count the number of times the infected agents meet the uninfected agents. We then show that this number is close to $M(t)$ by demonstrating that the number of overcounts is moderate, which yields the desired result. The device we use to count the number of meetings, however, is different from the one we used for Lemma 2.4.2. In Lemma 2.4.2, we couple each of the walks in \mathcal{V}^3 with their unbounded counterparts; since we only focus on a short time frame, the bounded walks largely coincide with the unbounded ones. Here, the right time frame to analyze is longer and the walks in \mathcal{V}^3 are more likely to hit the boundary. It becomes less helpful to relate these walks with the unbounded ones. Our analysis, instead, utilizes the mixing time property of \mathcal{V}^3 .

Specifically, we cut Δt into disjoint time intervals, each of which is of size $cn^2 \log n$ steps for some constant c to be determined later. We refer the k -th time interval as the k -th round. The total number of rounds in Δt steps is thus n/m .

We couple the diffusion process with a slower one. First, only agents in A^f are allowed to transmit the virus. An agent in A^u will not be able to infect others even if it becomes infected. This rule holds throughout the Δt time increment.

In each round, we also impose more specific constraints on the diffusion rule as follows. At the beginning of the k -th round (for any k), we first wait for $c_0 n^2 \log n$ steps so that the distribution of each agent is $1/(16n^3)$ -close to uniform distribution (see Definition 2.6.10 and Lemma 2.6.11 for details; c_0 is an appropriate constant that exists as a result of Lemma 2.6.11). Within these time steps, no agent becomes

infected even if it meets a previously infected agent. After these steps, for an arbitrary $a_i \in A^f$ and $a_j \in A^u$, let $X_{i,j}^k = 1$ if both of the following conditions hold:

- the L_1 -distance between a_i and a_j is between $n/450$ and $n/500$.
- the L_1 -distance between a_i and any boundary is at least $n/20$.

Since $c_0 n^2 \log n$ is already the mixing time for random walks on \mathcal{V}^d , it is straightforward to see that with $\Omega(1)$ probability $X_{i,j}^k = 1$, for any k .

After $c_0 n^2 \log n$ steps at k th round, our slower diffusion rule allows $a_i \in A^f$ to transmit its virus to $a_j \in A^u$ at the k th round only if

- $X_{i,j}^k = 1$.
- a_i meets a_j after the waiting stage and before the round ends.
- a_i and a_j have not visited any boundary after the waiting stage before they meet. In other words, an agent $a_i \in A^f$ ($a_j \in A^u$ resp.) loses its ability to transmit (receive resp.) the virus when it hits the boundary.

Let $Y_{i,j}^k$ be an indicator random variable that sets to 1 if and only if $a_i \in A^f$ transmits its virus to $a_j \in A^u$ under the slower diffusion rule at the k th round, *pretending that a_j is uninfected at the beginning of the k -th round even if it gets infected in the previous rounds*. Hence $Y_{i,j}^k$, for a specific i and j , can be 1 for more than one k . This apparently unnatural definition is used for the ease of counting in the sequel.

By Lemma 2.2.6,

$$\Pr[Y_{i,j}^k = 1] \geq \Pr[Y_{i,j}^k = 1 \mid X_{i,j}^k = 1] \Pr[X_{i,j}^k = 1] = \Omega(1/n).$$

Therefore, we have

$$\mathbb{E} \left[\sum_{i,j,k} Y_{i,j}^k \right] = \Omega \left(\frac{m_1 m_2}{m} \right) \geq \frac{\tau_1 m_1 m_2}{m}, \quad (2.28)$$

for some constant τ_1 .

We briefly lay out our subsequent analysis. We want to show two properties:

1. $\Pr[\sum_{i,j,k} Y_{i,j}^k = \Omega(\min\{m_1, m_2\})] = \tilde{\Omega}(1)$.
2. For all j , $\sum_{i,k} Y_{i,j}^k = \tilde{O}(1)$ with high probability.

We claim that these two properties together conclude our result. Roughly speaking, when $\left(\sum_{i,j,k} Y_{i,j}^k = \Omega(\min\{m_1, m_2\})\right)$ and $\left(\forall j : \sum_{i,k} Y_{i,j}^k = \tilde{O}(1)\right)$ occur, each $a_j \in A^u$ meets at most $\tilde{O}(1)$ agents in A^f while the total number of meetings between infected and uninfected agents is $\min\{m_1, m_2\}$. Consequently, the total number of uninfected agents that ever meet an infected agent is $\tilde{\Omega}(\min\{m_1, m_2\})$, hence our conclusion follows.

To prove the first property, we need to show with high probability, for any j , we have $\sum_{i,k} Y_{i,j}^k = \tilde{O}(1)$. Similarly, we also need to show with high probability, for any i , $\sum_{j,k} Y_{i,j}^k = \tilde{O}(1)$. Combining both of these we have $\sum_{i,j,k} Y_{i,j}^k = \tilde{O}(\min\{m_1, m_2\})$ with high probability. Together with Equation 2.28, some rearrangement of terms and Chernoff bounds, we can conclude that $\Pr[\sum_{i,j,k} Y_{i,j}^k = \tilde{\Omega}(\min\{m_1, m_2\})] = \tilde{\Omega}(1)$.

We now carry out this scheme. We proceed to show that

$$\Pr[\forall j : \sum_{i,k} Y_{i,j}^k = \tilde{O}(1)] \geq 1 - \exp(-\Omega(\log^2 n)). \quad (2.29)$$

and note that showing $\sum_{j,k} Y_{i,j}^k = \tilde{O}(1)$ can be done similarly. We prove Equation 2.29 via the following two steps:

1. first, we show that with high probability, $\sum_i Y_{i,j}^k = \tilde{O}(1)$ for any fixed k and j .
2. second, we show that with high probability, the number of k 's such that $\sum_i Y_{i,j}^k > 0$ is $\tilde{O}(1)$ for all j .

Intuitively, the first step ensures that there will not be too many meetings associated with a_j for any single round. The second step specifies an upper bound on the number of rounds in which a_j meets at least one infected agent. When both event occurs, the total number of meetings for a_j is $\tilde{O}(1)$.

Let us start with the first step. Fix a specific k and $a_j \in A^u$, by Corollary 2.2.5, we have

$$\begin{aligned}
 \Pr \left[\sum_{a_i \in A^f} Y_{i,j}^k \geq \log^2 n \mid X_{i,j}^k \right] &\leq \binom{\sum_{a_i \in A^f} X_{i,j}^k}{\log^2 n} \left(\frac{c_1 \log^2 n}{n} \right)^{\log^2 n} \\
 &\leq \binom{m_1}{\log^2 n} \left(\frac{c_1 \log^2 n}{n} \right)^{\log^2 n} \\
 &\leq \exp(-\log^2 n \log \log n).
 \end{aligned}$$

By a union bound, we can also conclude that

$$\Pr \left[\exists k \leq \frac{n}{m} : \sum_{a_i \in A^f} Y_{i,j}^k \geq \log^2 n \right] \leq \exp\left(-\frac{1}{2} \log^2 n \log \log n\right). \quad (2.30)$$

Next, let us move to the second step. Let us define a family of indicator random variables $I(j, k)$, which sets to 1 if and only if $\sum_{a_i \in A^f} Y_{i,j}^k \geq 1$. When j and k are fixed, we can compute the probability $\Pr[I(j, k) = 1]$:

$$\Pr[I(j, k) = 1] = \mathbb{E}[I(j, k)] \leq \mathbb{E} \left[\sum_{a_i \in A^f} Y_{i,j}^k \right] \leq \frac{\tau_1 m_1}{n}.$$

The probability holds regardless of the history of the process up to the time the k th round starts because $c_0 n^2 \log n$ time steps are used at k th round to shuffle the agents

so that they are distributed sufficiently uniform after these steps. We may apply a special case of Chernoff bound (see, e.g., Theorem A.1.2) to show that $\Pr[\sum_k I(j, k) > \log^2 n] < \exp(-\log^3 n)$.

Therefore, we have

$$\Pr \left[\exists a_j \in A^u : \sum_{k \leq \frac{n}{m}} I(j, k) > \log^2 n \right] \leq \exp(-\Theta(\log^2 n)). \quad (2.31)$$

For a specific $a_j \in A^u$, when both $(\sum_k I(j, k) < \log^2 n)$ and $(k : \sum_i Y_{i,j}^k \leq \log^2 n)$, we know that $\sum_{i \in A^f, k} Y_{i,j}^k \leq \log^4 n$. Hence Equation 2.30 and 2.31 imply $\Pr[\sum_{i \in A^f, k} Y_{i,j}^k > \log^4 n] \leq \exp(-\Theta(\log^2 n))$ and therefore

$$\Pr \left[\exists a_j \in A^u : \sum_{a_i \in A^f, k} Y_{i,j}^k > \log^4 n \right] \leq \exp(-\Theta(\log^2 n)) \quad (2.32)$$

Similarly, we can show

$$\Pr \left[\exists a_i \in A^f : \sum_{a_j \in A^u, k} Y_{i,j}^k > \log^4 n \right] \leq \exp(-\Theta(\log^2 n)) \quad (2.33)$$

Equation 2.32 and 2.33 yield

$$\Pr \left[\sum_{i,j,k} Y_{i,j}^k < \min\{m_1, m_2\} \log^4 n \right] \geq 1 - \exp(-\Theta(\log^2 n)). \quad (2.34)$$

This gives the first property in the discussion following Equation 2.28. Moreover, Equation 2.32 gives the second property.

Now, by using similar argument in the proof of Lemma 2.4.2, Equation 2.28 and 2.34 together give

$$\Pr \left[\sum_{i,j,k} Y_{i,j}^k \geq \frac{\tau_1 \min\{m_1, m_2\}}{4} \right] \geq \Pr \left[\sum_{i,j,k} Y_{i,j}^k \geq \frac{\tau_1 m_1 m_2}{2m} \right] \geq \log^{-5} n \quad (2.35)$$

When $\left(\sum_{i,j,k} Y_{i,j}^k \geq \frac{\tau_1 \min\{m_1, m_2\}}{4}\right)$ and $\left(\forall a_j \in A^u : \sum_{a_i \in A^f, k} Y_{i,j}^k \leq \log^4 n\right)$, the total number of infected agents is at least $\frac{\tau_1 \min\{m_1, m_2\}}{4 \log^4 n}$. Hence, by setting $c = 2c_0$, and using Equation 2.32 and 2.35, our lemma follows. \square

From this we can mimic the argument that appeared in Proposition 2.4.13 to reach the conclusion below:

Corollary 2.5.5. *Consider the diffusion process in which $m < \frac{n}{\log^2 n}$. Fix a specific time t , and let A^f and A^u be the set of infected and uninfected agents at t such that $|A^f| = m_1$ and $|A^u| = m_2$. Let $M(t)$ be the number of new infected agents between time t and time $t + \frac{n^3}{m} \log^{14} n$. Assume the agents are arbitrarily (in an adversarial manner) distributed at time t , we have*

$$\Pr \left[M(t) \geq \min \left\{ m_1, \frac{m_2}{2} \right\} \right] \geq 1 - \exp(-\log^2 n).$$

Similar to Lemma 2.4.13, Corollary 2.5.5 estimates the growth rate of infection as either doubling the number of infected agents or halving the uninfected ones within a certain time interval. One can then show that this implies Lemma 2.5.3. The argument is analogous to Chapter 2.4 and hence is skipped here.

2.6 Random walks properties

This section states some peripheral results for random walks that we used in our analysis.

Theorem 2.6.1. (First Passage Time, Chapter 3 of [42]) *Let $\{\mathbf{S}_t : t \in \mathbf{N}\}$ be a one dimensional random walk from the origin. The probability $\varphi_{r,t}$ that the first passage through r occurs at time t is given by*

$$\varphi_{r,t} = \frac{r}{t} \binom{t}{\frac{t+r}{2}} 2^{-t} \approx \sqrt{\frac{2}{\pi}} \frac{r}{\sqrt{t^3}} e^{-r^2/(2t)}.$$

Therefore, there exists constant C , such that for $r, t \geq C$, we have

$$\varphi_{r,t} \in \left(\frac{1}{2} \frac{r}{\sqrt{t^3}} e^{-r^2/(2t)}, \frac{r}{\sqrt{t^3}} e^{-r^2/(2t)} \right).$$

Lemma 2.6.2. *Let S be a bounded random walk in $[-n, n]$ starting from position P . For any other position $Q \in [-n, n]$ with $|P - Q| \geq \log^2 n$, the probability that S visits Q within $|P - Q|^2 / \log^4 n$ time steps is at most $\exp(-\log^3 n)$ when n is sufficiently large.*

Proof. Let us couple S with an unbounded random walk S' that also starts at P in the natural way, i.e. S and S' share the same random tosses to drive their moves.

First, we claim that at the first time S visits Q , the number of distinct lattice points S' visits is at least $|P - Q|$. This claim can be seen through analyzing the following two cases.

Case 1. The walk S never visits a boundary before its first visit to Q . In this case, S' coincides with S , which implies S' also visits all the lattice points between P and Q . The claim therefore follows.

Case 2. The walk S visits a boundary before it first visits Q . In this case, the boundary that S visits and the point Q lie on different sides of P . In other words, the distance between this boundary and Q is at least $|P - Q|$. Now let us only consider the time interval between the last time S visits the boundary (namely, t_0) and the first time S

visits Q . The trajectory of S' within this time interval is identical to the trajectory of S (up to an offset produced between time 0 and t_0). Therefore, from t_0 to the first time S visits Q , the coupled walk S' visits at least $|P - Q|$ distinct lattice points.

An immediate consequence of our claim is that a necessary condition for S to visit Q is that S' has to visit either $P - \frac{|P-Q|}{2}$ or $P + \frac{|P-Q|}{2}$. By Theorem 2.6.1, the probability S' ever visits either of these points within time $|P - Q|^2 / \log^4 n$ is at most $\exp(-\log^3 n)$ when n is sufficiently large, which completes our proof. \square

The next lemma concerns the first passage time for a random walk over bounded space.

Lemma 2.6.3. *Let S be a random walk on $\mathcal{V}^1 = \{-n, \dots, n\}$ that starts at A . Let B be a point on \mathcal{V}^1 such that $|B - A| = r$. Let T be the first time S visits B . Fix an arbitrary constant c , we have:*

$$\Pr[T \leq cr^2] = \Omega(1).$$

Proof. Without loss of generality, let us assume $-n \leq A \leq B \leq n$. We couple S with an unbounded random walk S' that also starts at A in the natural way, i.e. having S and S' share the same random tosses to drive their moves. Let T' be the first time S' visits B . We first show that $T' \geq T$. Note that before T' , S' is always to the left of B , and hence n . It is then easy to see that S is always overlapping or to the right of S' before T' . Hence S' hitting B at T' implies that S has already hit it at a time before or at T' .

Finally, by Theorem 2.6.1, we have

$$\Pr[T \leq cr^2] \geq \Pr[T' \leq cr^2] = \Omega(1).$$

□

Corollary 2.6.4. *Let S be a random walk on \mathcal{V}^1 that starts at A , and B be a point on \mathcal{V}^1 such that $|B - A| = r$. Let c_1 and c_2 be two arbitrary constants and let $t = c_1 r^2$. We have*

$$\Pr[|S_t - B| \leq c_2 r] = \Omega(1).$$

Proof. Let T be the first time S visits B . By Bayes' rule we have

$$\begin{aligned} & \Pr[|S_t - B| \leq c_2 r] \\ & \geq \Pr[|S_t - B| \leq c_2 r \mid T < t] \Pr[T < t] \\ & = \Pr[|S_t - S_T| \leq c_2 r \mid T < t] \Pr[T < t] \end{aligned}$$

By Lemma 2.6.3, $\Pr[T \leq t] = \Omega(1)$. Next we claim $\Pr[|S_t - S_T| \leq c_2 r \mid T < t] = \Omega(1)$. This can be seen by showing $\Pr[|S_t - S_T| \leq c_2 r \mid T] = \Omega(1)$ uniformly over $T \in [1, t]$. For this, note that $\Pr[|S_t - S_T| \leq c_2 r \mid T] = \Pr[|\tilde{S}_\tau| \leq c_2 r]$ where $\tau = t - T$ and \tilde{S} is a random walk starting at 0. We then write $\Pr[|\tilde{S}_\tau| \leq c_2 r] = \Pr[|\tilde{S}_\tau/\sqrt{\tau}| \leq c_2 r/\sqrt{\tau}] \geq \Pr[|\tilde{S}_\tau/\sqrt{\tau}| \leq c_2/\sqrt{c_1}] = \Omega(1)$ by Gaussian approximation on $\tilde{S}_\tau/\sqrt{\tau}$. Therefore, $\Pr[|S_t - B| \leq c_2 r] = \Omega(1)$. □

For a d -dimensional *unbounded* random walk starting from the origin, recall that $p_d(t, \vec{x})$ is the probability that the walk visits position \vec{x} at time t . Let $q_d(t, x)$ be the probability that the random walk visits \vec{x} *within* time t . When $d = 3$, we will silently drop the subscripts and write the functions as $p(\cdot, \cdot)$ and $q(\cdot, \cdot)$.

Theorem 2.6.5. [62] *The function $p_d(t, \vec{x})$ has the following analytic form, when $t - \|\vec{x}\|_1$ is even:*

$$p_d(t, \vec{x}) = \frac{2}{t^{d/2}} \left(\frac{d}{2\pi} \right)^{d/2} \exp \left\{ \frac{-d\|\vec{x}\|_2^2}{2t} \right\} + e_t(\vec{x}),$$

where $|e_t(\vec{x})| \leq \min \{O(t^{-(d+2)/2}), O(\|\vec{x}\|_2^{-2} t^{-d/2})\}$. $p_d(t, \vec{x}) = 0$ when $t - \|\vec{x}\|_1$ is odd.

Theorem 2.6.6. [4] *The function $q_d(t, \vec{x})$ satisfies the following asymptotic relations:*

- If $d = 2$, $\vec{x} \neq \vec{0}$, and $t \geq \|\vec{x}\|_2^2$, then we have

$$q_2(t, \vec{x}) = \Omega\left(\frac{1}{\log \|\vec{x}\|_2}\right).$$

- If $d \geq 3$, $\vec{x} \neq \vec{0}$, and $t \geq \|\vec{x}\|_2^2$, then we have

$$q_d(t, \vec{x}) = \Omega\left(\frac{1}{\|\vec{x}\|_2^{d-2}}\right).$$

When $d \geq 3$, it is not difficult to see that the above asymptotic result is tight by using Markov inequality:

Corollary 2.6.7. *When $d \geq 3$, the function $q_d(t, \vec{x})$ satisfies the following asymptotic relation for $t > \|\vec{x}\|_2^2$*

$$q_d(t, \vec{x}) = \Theta\left(\frac{1}{\|\vec{x}\|_2^{d-2}}\right).$$

Next we show for any random walk that could start near the boundary, waiting for a short period allows the walk to both stay away from the boundary and be sufficiently close to where it starts.

Lemma 2.6.8. *Consider a random walk S over the d -dimensional space \mathcal{V}^d that starts at \vec{x} , where $\vec{x} = (x_1, \dots, x_d)$ is an arbitrary point in the space. Let $\vec{c} = (c_1, \dots, c_d)$ be a point in \mathcal{V}^d such that $\|\vec{c} - \vec{x}\| = \Theta(r)$. Also let $t = r^2$. We have*

$$\Pr[S_t \in \mathbb{B}(\vec{c}, r)] = \Omega(1). \quad (2.36)$$

Proof. Recall that at each step, the random walk S uniformly selects a neighboring point to move to. We may also interpret a move of S as if it first randomly selects an axis to move along and next decides which one of the two directions to take when the axis is fixed. Let T_i be the number of the walk's move that are along the i -th axis within t steps. Define the event e as:

$$e = \left\{ \forall i : \frac{1}{2} \cdot \frac{t}{d} \leq T_i \leq \frac{5}{4} \cdot \frac{t}{d} \right\}.$$

By Chernoff bounds, we have for any specific $i \in [d]$,

$$\Pr \left[\frac{1}{2} \cdot \frac{t}{d} \leq T_i \leq \frac{5}{4} \cdot \frac{t}{d} \right] \geq 1 - \exp(-\Omega(t)) \geq 1 - \frac{1}{4d}$$

for sufficiently large t . Therefore,

$$\Pr[e] \geq 1 - d \cdot \frac{1}{4d} \geq \frac{3}{4}.$$

Let $(S_t)_i$ be the i -th coordinate of the point S_t . We next compute $\Pr[S_t \in \mathbb{B}(\vec{c}, r) \mid e]$:

$$\begin{aligned} & \Pr[S_t \in \mathbb{B}(\vec{c}, r) \mid e] \\ &= \mathbb{E} [\Pr[S_t \in \mathbb{B}(\vec{c}, r) \mid T_1, \dots, T_d, e] \mid e] \\ &= \mathbb{E} \left[\Pr \left[\bigwedge_{i \in [d]} (S_t)_i \in [c_i - r, c_i + r] \mid T_1, \dots, T_d, e \right] \mid e \right] \quad (\text{By the definition of } \mathbb{B}(\vec{c}, r)) \\ &= \mathbb{E} \left[\prod_{i \in [d]} \Pr \left[(S_t)_i \in [c_i - r, c_i + r] \mid T_i, e \right] \mid e \right] \end{aligned}$$

The last equality holds because the moves along the i -th axis are independent of the moves along other axes when T_i is known. Next, using Corollary 2.6.4, we have

$$\Pr \left[(S_t)_i \in [c_i - r, c_i + r] \mid T_i, e \right] = \Omega(1).$$

Therefore,

$$\begin{aligned} \Pr[S_t \in \mathbb{B}(\vec{c}, r) \mid e] &= \mathbb{E} \left[\prod_{i \in [d]} \Pr \left[(S_t)_i \in [c_i - r, c_i + r] \mid T_i, e \right] \mid e \right] \\ &= \mathbb{E} \left[\prod_{i \in [d]} \Omega(1) \mid e \right] = \Omega(1). \end{aligned}$$

Finally, we have

$$\Pr[S_t \in \mathbb{B}(\vec{c}, r)] \geq \Pr[S_t \in \mathbb{B}(\vec{c}, r) \mid e] \cdot \Pr[e] = \Omega(1).$$

□

Corollary 2.6.9. *Let r be sufficiently large and $r \leq \frac{n}{2(2\beta+6)}$, where β is an arbitrary constant between 1 and $80d$. Let $A = \vec{x}$ and B be two points in \mathcal{V}^d such that $\|A - B\|_1 \leq r$. Consider two bounded random walks S^1 and S^2 in \mathcal{V}^d that start with A and B respectively. Then, with $\Omega(1)$ probability, at time $t = r^2$,*

- S_t^1 is at least $\beta \cdot r$ away from any of the boundaries,
- $\|S_t^1 - A\|_\infty \leq (\beta + 2)r$, and
- $\|S_t^1 - S_t^2\|_\infty \in (r, 3r)$.

Proof. Let us first find an arbitrary $\vec{c} = (c_1, \dots, c_d)$ such that

- For all $i \in [d]$: $|c_i - x_i| = (\beta + 1)r$, i.e., $\|\vec{c} - \vec{x}\| = O(r)$.
- For all $i \in [d]$: $-n + (\beta + 1)r \leq c_i \leq n - (\beta + 1)r$, i.e., \vec{c} is sufficiently away from the boundary.

We set up β in a way that such \vec{c} always exists. By Lemma 2.6.8, we have $\Pr[S_t^1 \in \mathbb{B}(\vec{c}, r)] = \Omega(1)$. Next, in case $S_t^1 \in \mathbb{B}(\vec{c}, r)$, let $\vec{d}(S_t^1)$ be an arbitrary point such that

- $|d_i(S_t^1) - (S_t^1)_i| = 2r$
- the distance between $\vec{d}(S_t^1)$ and any boundary is at least βr .
- $(\beta + 1)r \leq \|\vec{d}(S_t^1) - B\|_\infty \leq (\beta + 5)r$.

Again by the way we designed β , such $\vec{d}(S_t^1)$ always exists so long as $S_t^1 \in \mathbb{B}(\vec{c}, r)$.

Using Lemma 2.6.8 again, we have

$$\Pr[S_t^2 \in \mathbb{B}(\vec{d}(S_t^1), r) \mid S_t^1 \in \mathbb{B}(\vec{c}, r)] = \Omega(1).$$

Therefore, we have

$$\Pr \left[\left(S_t^2 \in \mathbb{B}(\vec{d}(S_t^1), r) \right) \wedge \left(S_t^1 \in \mathbb{B}(\vec{c}, r) \right) \right] = \Omega(1).$$

Finally, observe that when $\left(S_t^2 \in \mathbb{B}(\vec{d}(S_t^1), r) \right) \wedge \left(S_t^1 \in \mathbb{B}(\vec{c}, r) \right)$, the three conditions specified in the Corollary are all met. This completes our proof. \square

2.6.1 Mixing time in graphs

Definition 2.6.10 (Statistical distance). *Let X and Y be two probability distributions over the same support \mathbf{P} . The statistical distance between X and Y is*

$$\Delta(X, Y) = \max_{T \subseteq \mathbf{P}} |\Pr[X \in T] - \Pr[Y \in T]|.$$

We also say that the distribution X is ϵ -close to Y if $\Delta(X, Y) = \epsilon$.

Lemma 2.6.11 (Mixing time for \mathcal{V}^3). *Consider a random walk that starts at point A for an arbitrary $A \in \mathcal{V}^3$. Let $\pi_t(A)$ be the distribution of the walk at time t , and π be the uniform distribution on the nodes in \mathcal{V}^3 . Let $\epsilon > 0$. When $t = \Theta(n^2 \log(1/\epsilon))$, we have*

$$\Delta(\pi_t(A), \pi) \leq \epsilon.$$

Although the mixing time of high dimensional torus were analyzed, we are not aware of any literature that pins down the exact mixing time for \mathcal{V}^3 . It is, however, straightforward to derive the mixing time in asymptotic form via computing the conductance of \mathcal{V}^1 (the one-dimensional grid) and using results on mixing times regarding tensoring graphs (e.g., Chapter 5 in [105] and [100]).

2.7 Existing techniques

This section briefly reviews existing lower bound and upper bound analysis techniques and explains the difficulties in generalizing them to the three dimensional case.

2.7.1 Lower bound

Two existing approaches that can potentially be adopted to our lower bound analysis are:

1. Geometrically understand the growth rate of the smallest ball that covers all the infected agents (hereafter, *the smallest covering ball*). An *upper bound* on the ball's growth rate translates into a lower bound on the completion time for diffusion. Examples of this approach include [5, 71].
2. Analyze the interaction of the agents locally to conclude that the influence of infection is constrained to a small region around the initially infected agent, over a small time increment. A union bound or recursive argument is then applied to give a global result. This approach is exemplified by [92].

Let us start with the first approach. Alves et al. and Kesten et al. [5, 71] assume the density of the agents is a constant; recall that the density of the agents is the ratio between the total number of agents and the volume of the space. Their model has infinite space, and hence there is no size parameter n . With this assumption, they obtain that the radius of the smallest covering ball grows linearly in time almost surely. Translating to our setting, an $o(1)$ density of agents would lead to a growth rate that is also linear in time t but scales in some way with the density. Directly applying a linear growth rate would still give a valid lower bound of order $\Omega(n)$ on the diffusion time, but this is substantially worse than the bound we need. One potential way to improve their argument is to analyze the scaling of the growth rate with respect to the density. While this approach may well be feasible, it is by no means immediate. For example, the analysis of [5, 71] appears to depend on the fact that two nearby agents have constant probability to meet within a small number of steps, which leads to the conclusion that uninfected agents near the smallest covering ball are quickly infected. This requires crucially that the density of agents is constant, and relaxing this assumption to $o(1)$ density appears non-trivial.

We have chosen instead to follow the technique developed by Pettarin et al. [92], extending it via our diffusion tree argument. We now argue that this extension appears necessary. Recall the island graph at time t defined in Definition 2.3.4. Pettarin et al.'s approach can be summarized by the following three steps:

1. At any time step, the island graph $G_t(\gamma)$ is constructed, where γ is an appropriately selected parameter.
2. Specify δt such that within δt time increment, w.h.p. a piece of virus is unable

to travel from one island to another.

3. Argue that the information has to travel across n/γ islands sequentially to complete the diffusion so that a lower bound $\frac{n}{\gamma} \cdot \delta t$ is established. The parameter n/γ is asymptotically optimal because the space \mathcal{V}^3 cannot pack more than n/γ islands along any directions (including those that are not parallel to the axes).

Now let us discuss the internal constraints over the parameters under this framework that prevents us from optimizing the lower bound for the 3-dimensional case.

At step 1, we need to decide γ . When γ is set to be larger than $n \cdot m^{-1/3}$ i.e. the critical percolation point [92], $G_t(\gamma)$ becomes connected w.h.p. and the subsequent arguments break down. Therefore, $\gamma \leq n \cdot m^{-1/3}$.

At step 2, for illustration let us only focus on two islands Isd_1 and Isd_2 , and let $a_1 \in \text{Isd}_1$ and $a_2 \in \text{Isd}_2$ be two arbitrary agents each from the two islands. We now need to decide on the value of δt . We are facing two options:

1. If δt is set to be smaller than γ^2 , then w.h.p. a_1 and a_2 do not meet in time δt [92].
2. If δt is larger than γ^2 , then with probability $\Theta(1/\gamma)$, a_1 and a_2 will meet in time δt (Lemma 2.2.3).

We consider both options to examine the quality of lower bounds we can get, using step 3 above. For the first option, the lower bound we get is $n\gamma \leq n^2 \cdot m^{-1/3}$, which is suboptimal. For instance when $m = n^{1.5}$, the lower bound is $n^{1.5}$ as opposed to $\tilde{\Omega}(n^{1.75})$. For the second option, regardless of the choice of δt , the lower bound

always *fails* to hold with probability $\Omega(1/\gamma) = \Omega(m^{1/3}/n)$ and so step 2 cannot be satisfied with high probability.

Our analysis corresponds to setting δt large, but doing a more careful analysis on the local infected region by considering a branching process that represents a historical trace of the infection. Our island diffusion rule is correspondingly modified from the rule of [92] to control the growth rate of this branching process.

2.7.2 Upper bound

We also explain why existing upper bound techniques such as those from [26, 92] do not appear to generalize immediately to the three dimensional case. The analyses in [26, 92], which are based on percolation, follow a proof strategy that contains two steps:

1. Let a_1 be the initially infected agent. Identify a ball \mathcal{B} (under L_∞ norm) of radius r that covers a_1 's initial position so that after t_1 time steps, where t_1 is a parameter to be decided, a constant portion of the agents in \mathcal{B} become infected (i.e. fraction of infected agents to total number of agents in \mathcal{B} is $\tilde{\Theta}(1)$). Moreover, these infected agents are well clustered i.e. at distance $\tilde{O}(r)$ from the ball \mathcal{B} .
2. Show that if a ball \mathcal{B}' has a constant portion of infected agents at time t , then at $t + t_2$, all adjacent balls with the same radius will also have a constant portion of infected agents. Here, t_2 is a parameter to be decided. Moreover, these newly infected agents are well clustered i.e. at distance $\tilde{O}(r)$ from the balls.

One usually also needs a good density condition i.e. agent density in any r -ball is $\Theta(m(r/n)^d)$. By repeatedly applying the second step, one can establish an upper bound on the time that all balls in \mathcal{V}^3 have constant portion of infected agents. Once this happens, usually it becomes straightforward to find the diffusion time. The asymptotic upper bound will be $\frac{n}{r} \cdot t_2 + t_1$.

Let us explain this in more detail for the case $d = 2$. Assume good density condition. First, we need to set $t_2 = \tilde{\Theta}(r^2)$ so that the newly infected agents at step 2 are well clustered. This ensures that the infected agents do not scatter uncontrollably outside a distance from the ball and jeopardize our next recursion. We now sketch a bound on r . Consider step 2. Suppose the number of infected agents in \mathcal{B}' at t is $m(r/n)^2 \times \tilde{\Theta}(1)$. By our choice $t_2 = \tilde{\Theta}(r^2)$, each infected agent in \mathcal{B}' has probability $\tilde{\Theta}(1)$ to meet each agent in the adjacent ball (by using Lemma 1 in [92]). Therefore, the expected number of infections in the adjacent ball is given by

$$\underbrace{m(r/n)^2 \times \tilde{\Theta}(1)}_{\text{\# of infected agents in } \mathcal{B}'} \times \underbrace{m(r/n)^2}_{\text{\# of uninfected agents in an adjacent ball}} \times \underbrace{\tilde{\Theta}(1)}_{\text{infection prob.}}$$

which, by the requirement of step 2, should be equal to $m(r/n)^2 \times \tilde{\Theta}(1)$. This gives $r = \Theta(\sqrt{n^2/m})$. Note that this also leads to the condition that the number of infected agents in \mathcal{B}' at t and the adjacent ball at $t + t_2$ are both $\tilde{\Theta}(1)$.

Now set $t_1 = \tilde{\Theta}(r^2)$ and so the number of infected agents in \mathcal{B} at time t_1 is $m(r/n)^2 \times \tilde{\Theta}(1) = \tilde{\Theta}(1)$. Note that both steps 1 and 2 are now satisfied. By recursively applying the second step, we can see that by time $\frac{n}{r} \cdot t_2 + t_1 = \tilde{\Theta}(n^2/\sqrt{m})$ all the balls in \mathcal{V}^2 will have $m(r/n)^2 \times \tilde{\Theta}(1)$ infected agents. Hence in the same order of time period $\tilde{\Theta}(n^2/\sqrt{m})$, all the agents in \mathcal{V}^2 will be infected. This time period gives the

optimal upper bound of the diffusion time for $d = 2$.

We now argue that this strategy does not work for $d = 3$. Let us attempt to mimic the above argument step by step. Again set $t_2 = \tilde{\Theta}(r^2)$ so that the infected agents are well clustered. Next, note that in contrast to the two-dimensional case, Lemma 2.2.3 states that the meeting probability of two random walks in \mathcal{V}^3 with initial distance r apart within time $\Theta(r^2)$ is $\Theta(1/r)$. Hence, in light of step 2, we require

$$\underbrace{m(r/n)^3 \times \tilde{\Theta}(1)}_{\text{\# of infected agents in } \mathcal{B}'} \times \underbrace{m(r/n)^3}_{\text{\# of uninfected agents in an adjacent ball}} \times \underbrace{\Theta(1/r)}_{\text{infection prob.}} = \underbrace{m(r/n)^3 \times \tilde{\Theta}(1)}_{\text{desired \# of infections}}$$

which gives $r = \tilde{\Theta}(\sqrt{n^3/m})$. Note that the number of infected agents in \mathcal{B}' at t and that of the adjacent balls at $t + t_2$ in step 2 are now both $m(r/n)^3 \times \tilde{\Theta}(1) = \tilde{\Theta}(\sqrt{n^3/m}) = \tilde{\Theta}(r)$.

We now try to set an appropriate value for t_1 . First, note that step 1 requires the number of infected agents in \mathcal{B} at time t_1 being $\tilde{\Theta}(r)$. Then the question is to find the approximate time for one initially infected agent to infect $\tilde{\Theta}(r)$ agents that are from \mathcal{B} . Moreover, we need that these infected agents do not travel at distance outside $\tilde{\Omega}(r)$ in the same time period.

To give a bound for this t_1 , let us look into the method of [92]. Note that in the case of $d = 2$, the number of agents in \mathcal{B} at any time is $\tilde{\Theta}(1)$. In this case, [92] suggests chopping the time t_1 into intervals each of length $\tilde{\Theta}(r^2)$. During each of these intervals, one only focuses on a pair of agents from \mathcal{B} and see if they meet each other; this method aims to reduce the analysis of correlation among multiple agents' meetings, a complicated quantity, to a sequence of independent problems that involve only the meeting of two random walks. Since there are only $\tilde{\Theta}(1)$ such pair

combinations, and that each such meeting probability is $\tilde{\Theta}(1)$, a $t_1 = \tilde{\Theta}(r^2)$ is enough to guarantee that the number of infected agents is $\tilde{\Theta}(1)$. Also these infected agents are well clustered at \mathcal{B} . Thus the argument works well for $d = 2$.

However, such an argument breaks down for $d = 3$ because now we are required to have $\tilde{\Theta}(r)$ infected agents at t_1 , and the meeting probability between any two agents is $\tilde{\Theta}(1/r)$. As a result the following tradeoffs cannot be balanced: 1) t_1 is set to be $\tilde{\Theta}(r^2)$ so that the infected agents are well clustered, but the number of infected agents at t_1 will only be $\tilde{\Theta}(1)$; 2) t_1 is set to be $\tilde{\omega}(r^2)$, but then the infected agents are not well clustered and may not constitute $\tilde{\Theta}(r)$ of infected agents within \mathcal{B} at t_1 . The first tradeoff appears if one uses the chopping argument of [92]: divide t_1 into intervals of length $\tilde{\Theta}(r^2)$. For each interval, observe the number of meetings between any infected and uninfected agents. This gives an expected total number of infections at t_1 as $r \cdot \tilde{\Theta}(1/r) = \tilde{\Theta}(1)$, which is less than the required number of $\tilde{\Theta}(r)$. Secondly, setting $t_1 = \tilde{\omega}(r^2)$ boosts up the number of infected agents, but also increases the chance that an infected agent escapes from the vicinity of \mathcal{B} . An accurate analysis of these two effects seems highly non-trivial and does not follow from the existing results of [92].

Finally, we mention the work of Clementi et al. [26] to deal with issues similar to above. At each step, conditioned on the positions of the infected agents, the infection event of each uninfected agent becomes independent of each other. The change in the infected population over time can then be analyzed. However, such analysis is possible in [26] because the agents in their model can jump at a distance $\Theta(\sqrt{n})$ at each step. This leads to much less serial dependence for each agent and consequently requires

less effort in keeping track of each agent's position. These phenomena, unfortunately, do not apply to our settings.

Chapter 3

Continuous distributed counting for non-monotonic streams

3.1 Background

A continuous distributed tracking model was introduced in [30] to address the challenges of designing an effective strategy to constantly track statistics in a dynamic, distributed environment. In this model, data arrive in multiple streams to a number of sites. All the sites are connected to a coordinator, and the goal of the coordinator is to continuously track some function of the aggregate data, and update it as new data arrives. An exact tracking would require each data sample to be communicated to the coordinator, which would incur a prohibitively large communication cost - linear in the size of the input stream. Similarly, space and time processing requirements may be very large. However, for most applications it is satisfactory to provide an approximate tracking. Thus, a general formulation of a continuous dis-

tributed tracking problem is to design an algorithm that will minimize the space, time and/or communication complexity while providing approximation guarantees on the tracking accuracy. Continuous distributed tracking problems have recently gained much interest in the research community [6, 31, 104, 112].

One of the basic building blocks for many of the existing algorithms is a counter. The goal of the counter is to report, with a given relative accuracy, the sum of values of all elements that have arrived across the aggregate stream arriving from distributed sites. The main assumption in almost all the previous works is that the input stream being counted is monotonic non-decreasing and, surprisingly, there is very little work on continuous distributed non-monotonic counters. Similarly, most of the previous algorithms using counters are not guaranteed to work correctly under non-monotonic input stream.

However, many data streams do not satisfy the monotonicity property. A simple motivating example is a voting/ranking application. Suppose users' votes come in a distributed stream. The goal is to keep a continuous track of which of the two options has a higher number of votes, and approximately by which voting margin. Here, the votes for each option can essentially be seen as two separate data streams, but we are interested in continuously monitoring the *difference* of the two streams, which is clearly non-monotonic. The naive approach of estimating the count of each option separately and then taking the difference will not provide a relative error guarantee for the difference.

Non-monotonic streams are common in many situations when dealing with instantaneous instead of cumulative phenomena, e.g. tracking a difference. One ex-

ample we analyze in more detail is monitoring a process that exhibits long-range dependency, a phenomena that has been found to be prevalent in nature, e.g. network traffic [74]. Also, non-monotonic counters are useful as building blocks in more complex algorithms whose inputs are not necessarily monotonic. A source of non-monotonicity could be the use of random projections that transform an input data stream into a non-monotonic stream. Another example that we discuss is a streaming implementation of a Bayesian linear regression problem (c.f. [14]), which is useful in the context of machine learning platforms for processing of large-scale data (e.g. [80]).

In this chapter we are interested in designing a continuous non-monotonic distributed counter with optimal communication complexity. We will also discuss its applications in different scenarios. We now define the problem in more detail.

3.1.1 Problem Definition

Consider a standard distributed streaming model where k sites are connected to a coordinator. Each site is allowed to communicate with the coordinator but they cannot communicate with each other directly (and a broadcast message counts as k messages). Data items a_1, \dots, a_n arrive at sites $\psi(1), \dots, \psi(n)$ respectively, at time instants $\tau_1 < \dots < \tau_n$. We shall refer to the item a_t as the t -th update. In a general continuous distributed monitoring problem, the coordinator is responsible to maintain a value of a function $f(a_1, \dots, a_t)$ at each time τ_t with a relative accuracy ϵ . We are interested in the *counter* problem, where the goal is to track a sum $S_t = \sum_{i \leq t} a_i$ of all the items that have arrived until time τ_t . The coordinator then needs to maintain an estimate that is between $(1 - \epsilon)S_t$ and $(1 + \epsilon)S_t$. Note that, by definition, the

counter problem has low space and time complexity, and thus we focus on minimizing communication complexity.

A monotonic counter only allows for positive increments. In particular, a canonical example of a monotonic counter [61] implies $a_t = 1$ for all t , meaning that a counter is incremented by one whenever an update arrives. We relax the assumption that a_t is positive, and we call this a *non-monotonic counting* problem. In the streaming literature, this input model is usually called a general (non-strict) turnstile model [87].

To the best of our knowledge, the only research so far dealing with non-monotonic input streams is Arackaparambil et al. [6], who studied the tracking of frequency moments F_p , where deletion operations are allowed ($a_t = 1$ denotes an insertion and $a_t = -1$ denotes a deletion). There, a strong negative result is established for the adversary input case for both tracking counts and tracking F_p when deletion is allowed: the worst-case communication complexity is $\Omega(n)$ messages for an input stream of n elements. It is straightforward to construct a worst-case input for the counter problem: consider the case where there is only one site and the updates consist of alternations between an insertion and a deletion. In this case, the true global counter evolves as the sequence $0, 1, 0, 1, \dots$. When one update is missed from the site, then the multiplicative error from the server becomes unbounded. Therefore, upon the arrival of each update, the site has to send a message to the server, which implies a communication lower bound of $\Omega(n)$ messages. While there is no way to circumvent this linear lower bound barrier for the worst-case input, it is natural to ask what the communication complexity is when the input is not fully adversarial

and consider the following question:

Can we design a continuous, distributed tracking protocols for counter for non-monotonic updates that has a sublinear communication complexity when the input is randomized?

In particular, we are interested in a random permutation model. In this model, an adversary first decides the entire sequence of updates a'_1, \dots, a'_n for all sites. We only assume that the sequence is bounded. Then, the “nature” decides a random permutation π . The final input to the sites is $a_1 = a'_{\pi(1)}, a_2 = a'_{\pi(2)}, \dots, a_n = a'_{\pi(n)}$. This model is very natural in large-scale settings (such as Internet scale, for example), where data is collected from a large number of individuals (e.g. Twitter or Facebook users). In such a model, a large amount of data is generated in short time intervals, and it is reasonable to assume that the order in which the individuals enter their inputs in the system is random, but the input itself can be arbitrary.

We are also interested if sublinear algorithms can be obtained for other types of random inputs that are well motivated by applications. For example, the use of random projections for computing sketches motivates to consider random i.i.d. updates. Another example is found in nature where many real-world phenomena exhibit self-similarity and long-range dependence (e.g. network traffic [74]) which where traditionally modeled by random processes such as fractional Brownian motion and found to be in good conformance with empirical data.

In all these data models, we shall assume that an adversary chooses the function $\psi(t)$ which defines how the stream is partitioned among the sites (an example is a load-balancing algorithm that can arbitrarily scatter inputs across sites). The adversary

can only decide the function $\psi(t)$ based on the information observed up to a point in time. This means that when the input is a random permutation, $\psi(t)$ can depend on the content of the updates decided by the adversary, the prefix of the permutation observed so far, and the values $\psi(1), \dots, \psi(t-1)$; while when the input is random, the function $\psi(t)$ can only depend on the values of a_1, \dots, a_{t-1} and the values of $\psi(1), \dots, \psi(t-1)$; . We will also assume that the times τ_1, \dots, τ_n at which the inputs arrive are decided by an adversary. This essentially implies that the coordinator and the other sites have no knowledge of the time instants at which an input arrives to its corresponding site, and any communication can only be initiated by a site that has received an update.

Note that our model is a strict generalization of the standard monotonic stream model for counting (c.f. Huang et al. [61]), where the updates are fixed to $a_t = 1$, and the arrival times and sites are adversarial. In our case, we relax the assumption on the value of updates and allow for randomly permuted adversarial or entirely random values of updates, while still keeping the adversarial data partitioning and arrival times.

3.1.2 Our Contributions

Our main results in this Chapter are matching upper and lower bounds on the communication complexity for a continuous, distributed, non-monotonic counter (up to a poly-logarithmic factor), which are sublinear in the size of the input. While these bounds hold for different types of inputs, we give a single algorithm that is optimal for all the types of inputs considered, and whose communication cost also matches

the corresponding lower bounds. The algorithm is lightweight in having only $\tilde{O}(1)$ space and update time complexity¹.

We first provide results for the case of Bernoulli i.i.d. input (Chapter 3.3) where we develop basic techniques that will be used in subsequent analysis. In the Bernoulli i.i.d. model, we assume that each update a_t is a Bernoulli random variable, with $\Pr[a_t = 1] = 1 - \Pr[a_t = -1] = p$, for some unknown parameter $p \in [0, 1]$. The counter value S_t is then a Bernoulli random walk with a drift $\mu = 2p - 1$. In the case of a Bernoulli i.i.d. input without a drift ($\mu = 0$), we show that a count can be tracked with $\tilde{O}(\sqrt{kn}/\epsilon)$ communication cost. In case of a Bernoulli i.i.d. input with an unknown drift $\mu \in [-1, 1]$, the achievable communication cost is $\tilde{O}(\min\{\sqrt{kn}, \sqrt{k}/|\mu|\}/\epsilon)$. In both cases, our algorithm does not need to know the drift. We also give matching lower bounds for most important cases (Chapter 3.4), showing the optimality of our algorithm.

This result should be compared with the communication cost $\tilde{\Theta}(\sqrt{k}/\epsilon)$ for a monotonic counter (with $a_t = 1$) that was recently established in [61]. We show that the same bound holds for a more general choice of updates (any i.i.d. Bernoulli input), as long as the drift is a positive constant. This is perhaps not entirely surprising, as for the constant drift case we use the algorithm from [61] as one of the building blocks for our algorithm. The key novel insight is that the communication cost increases to $\tilde{O}(\sqrt{kn}/\epsilon)$ when the drift is $|\mu| = O(1/\sqrt{n})$. Thus, we demonstrate that we are still able to track the count with a sublinear communication cost, and we describe the parameter ranges in which the cost is polynomial vs. polylog in the input size.

¹We use $\tilde{O}(x) = x \log^{O(1)}(nk/\epsilon)$ notation to ignore log factors.

We next turn to our main results for the permutation model (Chapter 3.3.3). Here we show that tracking is achievable with $\tilde{O}(\sqrt{kn}/\epsilon)$ communication cost and we give a matching lower bound (Chapter 3.4). This is to be contrasted with $\Theta(n)$ lower bound for a non-monotonic counter in a fully adversarial setting [6]. We show that, in a setting where all other parameters are chosen by an adversary, randomly permuting an arbitrary non-monotonic input is enough to permit a tracking algorithm with a sublinear communication cost. This shows that a sublinear tracking of non-monotonic input is still possible in a large number of real-world scenarios.

We further show that our algorithm can track a fractional Brownian motion with Hurst parameter $H \in [1/2, 1)$, where $1 < \delta \leq 1/H$ is an arbitrary parameter (Chapter 3.3.4) with total expected communication cost of $\tilde{O}(k^{\frac{3-\delta}{2}} n^{1-H}/\epsilon)$ messages. For the case of independent increments ($H = 1/2$), we get the same bound as before. For the case of positively correlated increments ($1/2 < H < 1$), which is of most interest in applications, we get a smaller communication cost. This is intuitive in view of the facts that increments are positively correlated which makes the process more predictable and the variance is larger. This in turn implies smaller expected residence of the count in the region of small values where there is a higher sensitivity to relative errors. Interestingly, the algorithm does not require to know the exact value of the parameter H , but only needs to have an estimate $1/\delta$ such that $H \leq 1/\delta$.

Finally, we show how our counter can be used as a building block for some instances of distributed tracking problems (Chapter 3.5). First, we construct an algorithm to track the second frequency moment (F_2 tracking) with $\tilde{O}(\sqrt{kn}/\epsilon^2)$ communication complexity (Chapter 3.5.1) and then provide a $\Omega(\min\{\sqrt{kn}/\epsilon, n\})$ lower

bound that is matching in both n and k . We also show how to use the non-monotonic counter as a building block for a Bayesian linear regression problem (Chapter 3.5.2), and show that the Bayesian linear regression can also be tracked with sublinear communication cost.

It is noteworthy that while the communication cost for non-monotonic random streams with subconstant drift is sublinear in the input size, this is significantly larger than for monotonic streams ($\tilde{O}(\sqrt{n})$ vs $\tilde{O}(1)$), which is because the problem is intrinsically more difficult. However, the fact that the communication cost is sublinear in the input size would still make the algorithm of appeal for practical applications. For example, Twitter users generate more than 10^8 tweets a day [102]. In this scenario, the communication cost of our algorithm for tracking a single counter would only be in the order of 10^4 messages per day, which is a significant reduction of the traffic load. Furthermore, our bounds are matching with the bounds for the monotonic counters in k and ϵ parameters.

Finally, we briefly discuss the main techniques used in this Chapter. As we are designing algorithms for random streams in a distributed environment, our solution naturally calls for an integration of different techniques from sampling theory, analysis of stochastic processes, classical streaming algorithms, and distributed algorithm design. The main ingredient in our algorithm to tackle an otherwise intractable problem in the adversarial setting is to make an optimal prediction on the evolution of the counter process using a scarce communication resource and adaptively changing the tracking strategy as we continuously update our predictions. Making the prediction requires us to understand the volatility structure of the counter process; in our

specific case, this boils down to the analysis of first passage time of random walks and random permutations. Designing a communication efficient tracking algorithm requires us to construct a sampling based protocol that can judiciously cope with the volatile structure of the process. To prove our matching lower bounds, we needed to carefully decompose the entire tracking process into disjoint segments so that we can apply results from the communication complexity and sampling theory separately on each segment and reach a strong lower bound that is polynomial in n .

3.1.3 Related Work

The research on functional monitoring in distributed systems has considered a variety of problems (e.g. [27, 29, 51, 52, 88]) including one-shot and continuous tracking query problems. To the best of our knowledge, Cormode et al. [30] is the first work that articulated the distributed computation model that we consider in the present Chapter. Substantial progress has recently been made on understanding various problems under this model, including drawing a sample with or without replacement (e.g. [31, 104]) and answering holistic queries such as tracking the rank of an item or computing a quantile (e.g. [28, 61, 112]).

The most closely related work to ours is the recent work of Huang et al. [61] and Arackaparambil et al. [6]. The work of Huang et al. examines the same counter problem as ours but assuming an important but more restrictive class of monotonic streams, where only positive increments are allowed. Our work relaxes this assumption on the input by allowing for non-monotonic streams where decrements are allowed (either i.i.d. or random permutation). Specifically, we assume that the rate of pos-

itive increments is $(1 + \mu)/2$, for some unknown drift parameter $-1 \leq \mu \leq 1$. For the special case of the drift parameter $\mu = 1$, our counter algorithm would solve the same counter problem as in [61] with the matching performance.

The work of Arackaparambil et al. considered non-monotonic functional monitoring in the adversarial setting, including the problem of continuously tracking F_2 that we study here. They established an $\Omega(n)$ lower bound for the 1-site case and an $\Omega(n/k)$ lower bound for the k -site case. For the random input stream that we study here, we establish a tight lower bound that is sublinear in n and grows with k , suggesting that our problem under random input may have a different structure than the one under fully adversarial input.

3.2 Algorithms and Notations

This section presents our algorithm for continuous distributed counting for non-monotonic streams. The algorithm is applicable to all input models, subject to choosing appropriate constants. In the subsequent sections we will show how to choose the constants for each input model under consideration.

In what follows, we shall write X_i be the i -th input (because the input is stochastic) and let $\mu = \mathbb{E}[X_i]$ be the *drift rate* of the counter process, \hat{S}_t be the coordinator's estimate of S_t . When the context is clear, we refer to a_t as both the t -th update and the item arrived at time t interchangeably though we shall be clear that the actual physical time is irrelevant in our algorithm. Also, we shall assume that each site always keeps track of the total number of updates arrived locally and maintain a local sum counter.

We will denote the sampling probability in SBC for the t -th update with $p_t = \text{Sample-Prob}(\hat{S}_t, t)$. For an algorithm, we define E_n to be the number of errors observed over an input of size n . We will be interested in algorithms such that $\Pr[E_n > 0] = O(1/n)$. We define M_n to be the number of messages transmitted over an input of size n . We note that it is sufficient to limit the size of a message to $O(\log n)$ bits to convey any possible counter value. Thus the number of bits transmitted over an input of size n is $\tilde{O}(M_n)$. We define R_t to be 1 if a message is sent to the coordinator and otherwise $R_t = 0$. We further denote with U_t the time until next message is sent to the coordinator as observed at time t . Similarly, we define V_t to be the time until the count process exits the ball $\mathbf{B}_\epsilon(S_t) = \{s \in \mathbb{Z} : |x - S_t| \leq \epsilon S_t\}$.

For the purpose of exposition, we will first start with the most fundamental case with Bernoulli i.i.d. increments. Recall that in this case $\Pr[X_t = 1] = p$ and $\Pr[X_t = -1] = 1 - p$. The expected increment in each step is then $\mu \triangleq p - (1 - p) = 2p - 1$. We shall refer to μ as the *drift* of the problem. We will first treat the case without drift ($\mu = 0$ and $p = 1/2$) and then the general case with an unknown drift. The analysis for other distributions heavily utilizes the idea developed for these simple cases.

We next introduce the basic constructs we need for the design of our distributed algorithm.

3.2.1 Building Blocks

The key novel building block in our scheme is:

Sampling and Broadcasting (SBC). In this protocol, the coordinator broadcasts its current estimate \hat{S}_t to all the sites at the beginning. Each site maintains a common

sampling rate $\approx 1/(\epsilon^2 \hat{S}_t^2)$ that depends only on the global estimate \hat{S}_t . Whenever a site receives a new update, it samples a Bernoulli random variable R_t with the above rate. If $R_t = 1$, the following actions will be carried out sequentially (invoking $\tilde{\Theta}(k)$ message exchanges):

1. The site signals the coordinator to sync all data.
2. The coordinator broadcasts a message to all the sites to collect their local counters.
3. Each site reports its local counter to the coordinator. The coordinator computes the new exact count and broadcasts this new count to all sites.

Upon receiving the new count \hat{S}_t , each site adjusts the sampling rate of the Bernoulli random variable to

$$\text{Sample-Prob}(\hat{S}_t, t) = \min \left\{ \frac{\alpha \log^\beta n}{\epsilon^2 \hat{S}_t^2}, 1 \right\} \quad (3.1)$$

where α and β are some appropriately chosen positive constants.

We will also use the following building blocks:

HYZ counter. In [61], a distributed counter is developed to track monotonic updates with a relative accuracy ϵ and error probability δ using $\tilde{O}(\frac{\sqrt{k}}{\epsilon} \log(1/\delta))$ communication when $k = O(1/\epsilon^2)$ and $\tilde{O}(k \log(1/\delta))$ communication when $k = \omega(1/\epsilon^2)$ (here, $\tilde{O}(\cdot)$ hides the poly-logarithmic dependencies on n). We shall refer this protocol as $\text{HYZ}(\epsilon, \delta)$.

Geometric Progression Search for μ (GPSearch). The goal of this building block is to produce a reliable estimator of μ . It will report an estimate $\hat{\mu}$ only when

sure w.h.p. that $\hat{\mu} \in [(1-\epsilon)\mu, (1+\epsilon)\mu]$, where ϵ is a given constant. It also guarantees that $\hat{\mu}$ is found before time $\Theta(\log n/\mu^2)$. We describe the GPSearch protocol in more details in Chapter 3.6.

Straightforward Synchronization (StraightSync). In this protocol, the coordinator pulls out the exact values of t and S_t from the sites in the beginning of the protocol. When a local site receives an update, it contacts the coordinator and executes the following steps sequentially:

1. The site sends both the total number of local updates and the local counter to the coordinator.
2. The coordinator updates the global count and the global number of updates.
3. The coordinator sends the updated global count and global number of updates to the site.

3.2.2 Algorithm Overview

Our algorithm, called **Non-monotonic Counter**, consists of two phases.

Phase 1: The first phase covers updates from $t = 0$ to $t = \tau$, where $\tau = c \log n/(\mu^2\epsilon)$ for some sufficiently large constant $c > 0$. During this phase, we have two communication patterns:

- When $(\epsilon\hat{S}_t)^2 \geq k$, we use the SBC protocol.
- When $(\epsilon\hat{S}_t)^2 < k$, we use the StraightSync protocol.

The coordinator shall make a broadcast when the algorithm makes a switch between SBC and StraightSync protocol.

Phase 2: The second phase covers from $t = \tau$ to $t = n$ (the second phase could be empty when $\tau \geq n$). In the second phase, the algorithm maintains a $\text{HYZ}(\Theta(\epsilon\mu), \Theta(1/n^2))$ to track the total number of positive updates and another $\text{HYZ}(\Theta(\epsilon\mu), \Theta(1/n^2))$ to track the total number of negative updates. The difference between the positive updates and the negative updates is the estimator maintained by the coordinator.

In addition, our GPSearch procedure is executed in the background, and it will be able to tell us a good estimate of μ , and decide when Phase 1 ends. When the algorithm changes from the first phase to the second phase, the coordinator shall make a broadcast to inform different sites of the phase change.

3.3 Upper Bounds

We now analyze *Non-monotonic Counter* for i.i.d. input, randomly ordered streams, and fractional Brownian motion. Recall that because the input is of stochastic nature, we shall write the updates as X_1, X_2, \dots, X_n instead of a_1, \dots, a_n to emphasize the randomness. Our analysis starts with the simplest case, where $k = 1$ and the input is i.i.d. with $\mu = 0$. Then we move to the more complex scenarios, in which there are multiple sites and unknown μ . Finally, we generalize our algorithms and analysis to the randomly ordered stream and fractional Brownian motion case, where the updates are no longer independent. Along the way, we shall explain the reasons

why we design the algorithm in such a way and gradually unfold the key techniques used in the analysis.

3.3.1 I.I.D. Input with Zero Drift

Recall the algorithm Non-monotonic Counter that is described in Chapter 3.2. To analyze the behavior of our algorithm we start by giving an upper bound for the single-site case ($k = 1$), and we then turn to the multi-site case. In the single-site case, we need to introduce a small modification to the algorithm. Since the site is aware of the exact counter value $\hat{S}_t = S_t$, there is no need for the straightforward stage and we assume that the algorithm is always in the broadcast stage. Also, there is no need for the coordinator to send messages back to the site.

We will use the sampling probability as defined earlier in (3.1) with $\alpha > 9/2$ and $\beta = 2$. The parameter α controls the tradeoff between communication complexity and the success probability. The choice of p_t is intuitive because the smaller S_t is, the more likely that a small change of S_t will cause large multiplicative change; therefore, the site should report the value to the coordinator with higher probability.

We have the following theorem:

Theorem 3.3.1. *For the single-site case, the randomized algorithm Non-monotonic Counter with the sampling probability as in (3.1), with $\alpha > 9/2$ and $\beta = 2$, guarantees to track the count within the relative accuracy $\epsilon > 0$ with probability $1 - O(1/n)$ and uses the total expected communication of $O(\min\{\sqrt{n}/\epsilon \cdot \log n, n\})$ messages.*

Before we present the full analysis, we comment on the intuition of using a sampling based algorithm and setting the sampling probability as specified in (3.1).

We want the site to send messages to the coordinator as infrequently as possible. Suppose that at time t , we have $S_t = s$ and a message is sent to the server. We need to understand what the next appropriate time would be to send another message. Ideally, this shall happen at the time where S_t first passes through either $s/(1 + \epsilon)$ or $s/(1 - \epsilon)$. Implementing this strategy is feasible when there is only one site but it is unclear how it can scale up to k site case (because the challenge of distributed tracking algorithms exactly lies in the difficulties of *exactly* tracing the aggregate statistics and thus it is hard to spot the exact first passage time). It is therefore desirable to use a “smooth” strategy by a site, i.e. the algorithm does not critically rely on the knowledge on the time when S_t first passes through some pre-specified points. The sampling based algorithm possesses such a property. We also need to estimate the sampling rate of the algorithm. Intuitively, it takes an unbiased random walk approximately $(\epsilon s)^2$ time to travel for a distance of about length ϵs (to hit either $s/(1 - \epsilon)$ or $s/(1 + \epsilon)$). When ϵs becomes sufficiently large, we even have a concentration result, i.e. with high probability, the time it takes to hit either $s/(1 - \epsilon)$ or $s/(1 + \epsilon)$ is $\tilde{\Theta}((\epsilon s)^2)$. Therefore, sampling at rate $\tilde{\Theta}(1/(\epsilon s)^2)$ is not only sufficient to maintain high accuracy but also optimal.

Proof of Theorem 3.3.1. We shall first analyze the communication cost of the algorithm. Then we shall argue our algorithm is correct with high probability.

Communication cost. We first show that the expected number of communicated messages is bounded as asserted. Let $\vartheta = \sqrt{\alpha}/\epsilon \cdot \log n$ and note

$$\mathbb{E}[R_t] = \Pr[|S_t| \leq \vartheta] + \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right].$$

Since $\Pr[|S_t| \leq \vartheta] = \Theta(\vartheta/\sqrt{t})$ and $\mathbb{E}[\frac{1}{S_t^2} I(|S_t| > \vartheta)] = \Theta(1/(\vartheta\sqrt{t}))$, it follows

$$\mathbb{E}[R_t] = \Theta(\vartheta/\sqrt{t}).$$

Hence, the expected number of transmitted messages is

$$\sum_{t \leq n} \mathbb{E}[R_t] = O(\vartheta\sqrt{n}) = O(\sqrt{n}/\epsilon \cdot \log n).$$

Correctness. We next establish the asserted bound on the probability of error. Let us write \mathcal{F}_t to be the σ -algebra generated by X_1, \dots, X_t and R_1, \dots, R_t , i.e. all the information available up to the t -th update is measurable by \mathcal{F}_t . Define the indicator variable R_t that sets to 1 if and only if at the t -th update the site sends a message to the coordinator. Notice that our algorithm guarantees that $R_1 = 1$. Let U_t be the number of updates until the next report is sent to the coordinator as observed at the t -th update, i.e. $U_t = \min\{\tau > 0 : R_{t+\tau} = 1\}$. We remark that U_t depends on a future event and thus, it is not measurable by \mathcal{F}_t . Next, let $V_t = \min\{\tau > 0 : S_{t+\tau} \notin \mathbf{B}_\epsilon(S_t)\}$ be the number of updates until the first instance at which the coordinator fails to track the counter within the relative accuracy ϵ , and let E_n be the number of such update instances. Notice that a necessary and sufficient condition that at least one error happens is that there exists at least one $t \leq n$ such that $R_t = 1$ and $V_t < U_t$. We thus have

$$\Pr[E_n > 0] = \Pr[R_t = 1 \text{ and } V_t < U_t, \text{ for some } 1 \leq t \leq n],$$

where $I(\cdot)$ is an indicator function that sets to 1 if and only if its parameter is true.

By using the union bound, we have

$$\Pr[E_n > 0] \leq \sum_{t \leq n} \mathbb{E}[R_t \cdot I(V_t < U_t)]. \quad (3.2)$$

Using the fact that R_t is measurable by \mathcal{F}_t , we have

$$\begin{aligned}
\mathbb{E}[R_t \cdot I(V_t < U_t)] &= \mathbb{E}_{\mathcal{F}_t}[\mathbb{E}[R_t I(V_t < U_t) \mid \mathcal{F}_t]] \\
&= \mathbb{E}_{\mathcal{F}_t}[R_t \mathbb{E}[I(V_t < U_t) \mid \mathcal{F}_t]] \\
&= \mathbb{E}_{\mathcal{F}_t}[R_t \Pr[V_t < U_t \mid S_t]] \\
&\leq \mathbb{E}_{\mathcal{F}_t}\left[R_t \cdot \max_s \Pr[V_t < U_t \mid S_t = s]\right] \\
&= \mathbb{E}_{\mathcal{F}_t}[R_t] \cdot \max_s \Pr[V_t < U_t \mid S_t = s].
\end{aligned}$$

We next proceed to give an upper bound for $\Pr[V_t < U_t \mid S_t = s]$. Note that for every $r \geq 0$, it holds

$$\begin{aligned}
\Pr[V_t < U_t \mid S_t = s] &= \Pr[V_t < U_t, V_t > r \mid S_t = s] + \Pr[V_t < U_t, V_t \leq r \mid S_t = s] \\
&\leq \Pr[r < U_t \mid S_t = s, V_t > r] + \Pr[V_t \leq r \mid S_t = s] \tag{3.3}
\end{aligned}$$

We start by giving a bound on $\Pr[V_t \leq r \mid S_t = s]$. Notice that under $S_t = s$ the distribution of V_t is equal to the distribution of the first passage time of either value $\lceil s/(1 - \epsilon) \rceil - s$ or value $\lfloor s/(1 + \epsilon) \rfloor - s$ for a symmetric random walk started at the origin. The following lemma follows by standard results from the theory of random walks (c.f. [73]) and the Hoeffding bound:

Lemma 3.3.2. *For every $r \geq 0$, it holds*

$$\Pr[V_t \leq r \mid S_t = s] \leq 2 \exp\left(-\frac{\left(\frac{\epsilon}{1-\epsilon}\right)^2 s^2}{2r}\right). \tag{3.4}$$

Proof. Let V_t^+ denote the number of steps until the random walk up-crosses the value $\lceil \frac{s}{1-\epsilon} \rceil$, starting from value s . Similarly, we define V_t^- to be the number of steps until

the random walk down-crosses the value $\lfloor \frac{s}{1+\epsilon} \rfloor$ starting from value s . Then,

$$\begin{aligned} \Pr[V_t \leq r | S_t = s] &\leq \Pr[V_t^+ \leq r | S_t = s] + \Pr[V_t^- \leq r | S_t = s] \\ &\leq 2 \Pr[V_t^+ \leq r | S_t = s]. \end{aligned}$$

Now, let $b = \lceil \frac{s}{1-\epsilon} \rceil - s$ and note that by the reflection principle of random walks, we have

$$\begin{aligned} \Pr[V_t^+ \leq r | S_t = s] &= \Pr[X_1 + X_2 + \dots + X_r = b] + 2 \Pr[X_1 + X_2 + \dots + X_r > b] \\ &\leq 2 \Pr[X_1 + X_2 + \dots + X_r \geq b]. \end{aligned}$$

By applying the Hoeffding's inequality, we bound the the probability in the right-hand side with $\exp(-\frac{b^2}{2r})$ which yields the asserted result. \square

From (3.4), we observe that $\Pr[V_t \leq r | S_t = s] \leq 2/n^c$ for given $c > 0$, iff it holds

$$(C1): \quad r \leq \frac{1}{2c \log n} \left(\frac{\epsilon}{1-\epsilon} \right)^2 s^2.$$

We next note

$$\Pr[r < U_t | S_t = s, V_t > r] \leq (1 - \rho_\epsilon(s))^r \quad (3.5)$$

where $\rho_\epsilon(s) = \text{Sample-Prob}(s/(1-\epsilon), t)$. Requiring that the right-hand side in the above inequality is less than or equal to $1/n^c$, we obtain

$$(C2): \quad \rho_\epsilon(s) \geq 1 - \exp\left(-\frac{c \log n}{(1-\epsilon)^2 r}\right).$$

Indeed, both conditions (C1) and (C2) hold true by taking $r = \frac{1}{2c \log n} \left(\frac{\epsilon}{1-\epsilon} \right)^2 s^2$ and $\rho_\epsilon(s) = \min\{\frac{2c^2 \log^2 n}{(\epsilon s)^2}, 1\}$. The latter choice is a sufficient condition for (C2) in view

of the fact that $\min\{x, 1\} \geq 1 - e^{-x}$, for $x \geq 0$. Therefore, we showed that for $\Pr[V_t < U_t | S_t = s] \leq 3/n^c$ to hold, it suffices that the sampling probability satisfies

$$p_t \geq \min\left\{\frac{2c^2 \log n}{\epsilon^2 S_t^2}, 1\right\}. \quad (3.6)$$

Combining with (3.2), we have

$$\begin{aligned} \Pr[E_n > 0] &\leq \sum_{t \leq n} \mathbb{E}[R_t] \cdot O(1/n^c) \\ &= \Theta(\vartheta n^{1/2-c}) = \Theta(n^{1/2-c} \log n). \end{aligned}$$

From the last inequality, we note that no error occurs with high probability provided that $c > 3/2$. Hence, in view of the inequality (3.6), it suffices to choose the sampling probability as in (3.1) with $\alpha = 2c^2 > 9/2$ and $\beta = 2$. This completes the proof. □

We now extend to the multiple site case. Let us first go through the intuition why we want to distinguish the two stages of the algorithm, the straightforward (StraightSync) stage and the broadcast (SBC) stage, described in Chapter 3.2. The main idea of our distributed algorithm is to simulate the behavior of the sampling algorithm for the single site case (Theorem 3.3.1). For that we require that each site has a good estimate \hat{S}_t of the global count S_t . As \hat{S}_t gets updated, the copies of the counter at all sites need to be updated, in order to maintain the correct local sampling rate. The only way to achieve so is to broadcast the counter, which would result in $\tilde{\Theta}(k)$ messages exchanged. The crucial observation here is that when \hat{S}_t gets updated frequently (i.e., when S_t is sufficiently small), broadcasting messages after each update could be wasteful. It may be even worse than the trivial approach where

a site synchronizes only with the coordinator whenever it receives an update (resulting in $\tilde{O}(1)$ messages). This “trivial” approach is captured in our straightforward strategy, and we switch to it whenever it is less expensive. Note that we can use the estimator \hat{S}_t instead of the actual value S_t to decide whether the broadcasting or the straightforward strategy has the smaller communication cost, because we guarantee a sufficiently high accuracy of the estimator.

We have the following theorem.

Theorem 3.3.3. *The randomized algorithm `Non-monotonicCounter` with the sampling probability as in (3.1), with α large enough positive constant and $\beta = 2$, guarantees to track the count within the relative accuracy $\epsilon > 0$ with probability $1 - O(1/n)$ and uses the total expected communication of $O(\min\{\sqrt{kn}/\epsilon \cdot \log n, n\})$ messages.*

Proof of Theorem 3.3.3. We need again to show that the algorithm is correct and the communication complexity is as described. We start with showing the correctness part.

Correctness. We will invoke a coupling argument that will allow us to reuse the results of Theorem 3.3.1. We couple the proposed multiple sites algorithm with the single site sampling algorithm with a different set of error parameters over the same set of input. Specifically, we also execute a single site algorithm with relative accuracy $\epsilon/3$ and success rate $1 - O(1/n^2)^2$, in parallel to the multiple sites algorithm over the same input sequence. We shall couple the random tosses in these two algorithms and show that when the single site algorithm makes no error, our multiple sites algorithm

²To boost the success rate, we need to use a larger constant in the sampling parameter, i.e. $\text{Sample-Prob}(S_t, t) = \min\left\{\frac{2(1+c)^2 \log^2 n}{\epsilon^2 \hat{S}_t^2}, 1\right\}$, any $c > 3/2$

will also make no error.

We need a few more notations. Let $p_{s,i}$ be the sampling rate for the single site algorithm and $R_{s,i}$ be its corresponding Bernoulli random variable. Let $p_{m,i}$ be the sampling rate for the multiple sites algorithm and $R_{m,i}$ be its corresponding Bernoulli random variable. When we are in the straightforward stage, we shall assume $p_{m,i} = 1$. Finally, let $\hat{S}_{s,t}$ be the estimator of the single site algorithm at time t and $\hat{S}_{m,t}$ be the estimator for the multiple sites algorithm.

We couple the Bernoulli random variable in the following way.

- When $p_{s,i} > p_{m,i}$: the two Bernoulli variables are sampled independently.
- When $p_{s,i} \leq p_{m,i}$: if $R_{s,i} = 1$, then we set $R_{m,i} = 1$; otherwise, we set $R_{m,i} = 1$ with probability $(p_{m,i} - p_{s,i})/(1 - p_{s,i})$ and $R_{m,i} = 0$ otherwise. One may see that we still have $\Pr[R_{m,i} = 1] = p_{m,i}$.

Now using the above coupling rules, we show that when the single site makes no error, our multiple sites algorithm also makes no error. Suppose on the contrary that at time t the multiple sites algorithm makes the first error. Then our algorithm ensures that for every $\tau < t$, it holds $p_{m,\tau} \geq p_{s,\tau}$ (by our choice of sampling probabilities), i.e. the multiple sites algorithm samples more frequently than the single site algorithm. Therefore, our coupling rule gives us $\hat{S}_{m,t} = S_{t_1}$ and $\hat{S}_{s,t} = S_{t_2}$, where $t_1 > t_2$, i.e. the multiple sites algorithm is holding a more recent value of the count. We can now get a contradiction using the following arguments,

1. At time t_2 , the single site algorithm's estimator is S_{t_1} and is correct. Therefore,

$S_{t_1} \in \mathbf{B}_{\epsilon_3}(S_{t_2})$, i.e.

$$|S_{t_2} - S_{t_1}| \leq \frac{\epsilon}{3}|S_{t_2}|. \quad (3.7)$$

2. At time t , the multiple site algorithm is wrong. Therefore, $S_{t_1} \notin \mathbf{B}_\epsilon(S_t)$, i.e.

$$|S_{t_1} - S_t| > \epsilon|S_t|. \quad (3.8)$$

3. At time t , the single site algorithm is correct, i.e. $S_{t_2} \in \mathbf{B}_\epsilon(S_t)$. We have

$|S_{t_2} - S_t| \leq \epsilon|S_t|$. We can use $\epsilon \leq 1$ to relax this inequality and get

$$|S_{t_2}| \leq 2|S_t|. \quad (3.9)$$

Using (3.7) and (3.8) and a triangle inequality, we have

$$|S_{t_2} - S_t| > \epsilon|S_t| - \frac{\epsilon}{3}|S_{t_2}| \geq \epsilon|S_t| - \frac{2\epsilon}{3}|S_t| \geq \frac{\epsilon}{3}|S_t|. \quad (3.10)$$

The second inequality holds because of (3.9). (3.10) implies that the single site algorithm errs at time t , which contradicts with our assumption.

Communication cost. We have the following types of communications,

1. At the straightforward stage, whenever there is an update, $O(1)$ messages are exchanged.
2. At the broadcast stage, whenever there is an update, $O(k)$ messages are exchanged.
3. At the beginning and the end of the broadcasting stage, the coordinator needs to make a broadcast to signal the stage change, which takes $\Theta(k)$ messages.

Notice that in order to change from the broadcast stage to straightforward stage, type 2 messages are sent for at least once. Therefore, the total complexity of the type 3 messages is asymptotically smaller than type 2 messages. We need to only focus on the communication complexity for the first two type of messages.

Let C_t be the communication cost associated with the t -th update and let $R_{m,t}$ indicates the event that a message is sent to the communicator after the t -th update ($R_{m,t}$ shall correspond with R_t in Theorem 3.3.1). Therefore, when $(\epsilon \hat{S}_t)^2 < k$, $C_t = 1$; otherwise, $E[C_t] = kE[R_{m,t}]$. We estimate C_t using the following rule:

- If $(1 - \epsilon)(\epsilon S_t)^2 \leq k$, we set $C_t = 1$;
- If $(1 + \epsilon)(\epsilon S_t)^2 > k$, we set $C_t = kE[R_{m,t}]$.

This rule intuitively gives a conservative guess on which stage we are in (conditioned on the estimator being correct). Notice that when $(1 - \epsilon)(\epsilon S_t)^2 < k < (1 + \epsilon)(\epsilon S_t)^2$, in this case, we can set $C_t = 1 + kE[R_{m,t}]$ without impacting the asymptotic behavior. The case where our estimator makes an error (and thus the above rules may not give an overestimate of C_t) is an asymptotically smaller term.

We next proceed with computing the expectation of C_t using our overestimation rule,

$$E[C_t] \leq \underbrace{\Pr[S_t \leq \frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}}]}_{\text{straightforward stage}} + \underbrace{kE[R_{m,t}I(S_t \geq \frac{\sqrt{k}}{\epsilon\sqrt{1+\epsilon}})]}_{\text{broadcast stage}} + \underbrace{O(1/n^2)}_{\text{estimator fails}} = O(\frac{\sqrt{k} \cdot \log n}{\epsilon\sqrt{t}}). \quad (3.11)$$

We can compute the above terms using Theorem 3.3.1. Thus, the total communication cost in expectation is $O(\sqrt{nk}/\epsilon \cdot \log^2 n)$. \square

3.3.2 I.I.D. Input with Unknown Drift

In the previous section we have seen that the communication complexity in the case with no drift is $\tilde{O}(\sqrt{n})$. However, the monotonic counter from [61] is a special case of our model with $\mu = 1$, and its communication complexity is $\tilde{O}(1)$. Clearly, we conclude that a positive drift might help. The natural question then is whether this observation holds for an arbitrary drift $\mu \neq 0$, and how can we exploit it when the drift is unknown.

To gain an intuition on the input's behavior for an arbitrary drift, it is helpful to re-parameterize each input X_t as $X_t = \mu + Z_t$, where μ is the drift term and Z_t is a random variable representing the “noise” term. We shall intuitively view Z_t as noise that behaves similar to Gaussian noise. We want to identify which term contributes more to the estimation error. Suppose $S_t = s$. It takes the drifting term ϵt time units to reach $\pm \epsilon s$ while it takes the noise term $(\epsilon s)^2$ to do so. When $\epsilon t < (\epsilon s)^2$, the drifting term dominates the process, otherwise, the noise term dominates the process. Approximating s by its mean $s \approx t\mu$ and solving the equation, $\epsilon t < (\epsilon s)^2$, we get $t \approx 1/(\mu^2 \epsilon)$. Therefore, the random walk S_t qualitatively behaves as follows: up to time $t = \Theta(1/(\epsilon \mu^2))$, the “noise” sum term dominates the process; after time $\Theta(1/(\epsilon \mu^2))$, the drifting term dominates the process. Therefore, intuitively, for $t \leq 1/(\epsilon \mu^2)$ we should use the algorithm that deals with the non-drift case, and for $t \geq 1/(\epsilon \mu^2)$ we might be able to use the monotonic counter HYZ.

Note that the algorithm does not know the actual value of the drift μ . We use an online estimator (the GPSearch algorithm, described in Chapter 3.2) to obtain an estimate $\hat{\mu}$. Our estimator is conservative in the sense that it does not report $\hat{\mu}$

until confident that it is within $[(1 - \epsilon')\mu, (1 + \epsilon')\mu]$ (the performance of the GPSearch estimator is discussed in Chapter 3.6). Once the estimator $\hat{\mu}$ is reported, we can safely switch to the monotonic counter HYZ.

However, we need to guarantee correctness of the algorithm even before we have an estimate of $\hat{\mu}$. The monotonic counter HYZ essentially samples with sampling probability $\Theta(1/(\epsilon t))$. So to guarantee the correctness before we know whether we are in the no-drift phase or in the drift phase, we need to sample with the maximum of the sampling rate $\Theta(1/(\epsilon^2 s^2))$ of the no-drift phase and the sampling rate $\Theta(1/(\epsilon t))$ of the monotonic counter. We shall choose a slightly more conservative rate by tuning the constants in the sampling probability (3.1) so that $\text{Sample-Prob}(S_t, t) \geq \tilde{\Theta}(1/\epsilon^2 s^2 + 1/\epsilon t)$ for all $t < 1/(\mu^2 \epsilon)$.

The crucial observation here is that this conservative way of sampling will not result in substantial increase in communication resource. Indeed, we have two types of unnecessary communication costs:

- Type 1: when $t \leq 1/(\epsilon \mu^2)$, the term $\tilde{\Theta}(1/\epsilon t)$ in the sampling rate is wasteful.
- Type 2: when $t > 1/(\epsilon \mu^2)$, the term $\tilde{\Theta}(1/(\epsilon s)^2)$ in the sampling rate is wasteful.

The total expected communication cost of type 1 is $O(\sum_{t \leq n} (1/t)) = O(\log n)$, which is acceptable. Computing the waste of type 2 is a bit more tedious, but we would be able to see that in expectation $\sum_{t \leq 1/(\epsilon \mu^2)} 1/(\epsilon^2 S_t^2) = \Omega(\sum_{t \leq n} 1/(\epsilon^2 S_t^2))$. In other words, $\sum_{1/(\epsilon \mu^2) < t \leq n} 1/(\epsilon^2 S_t^2) = O\left(\sum_{t \leq 1/(\epsilon \mu^2)} 1/(\epsilon^2 S_t^2)\right)$, i.e the total wasted communication from the term $\tilde{\Theta}(1/(\epsilon s)^2)$ is bounded by the total “useful” communication from the same term. Therefore, the conservative sampling strategy is also optimal.

In what follows, we shall mimic the analysis Chapter 3.3.1 and study the single site case and multiple site case separately.

Single site case. For the sake of our analysis, we shall analyze the performance of our algorithm with the following sampling probability (for a constant $\alpha > 0$):

$$\text{Sample-Prob}(S_t, t) = \min \left\{ \frac{\alpha \log^2 n}{\epsilon^2 S_t^2} + \frac{\alpha \log^3 n}{\epsilon t}, 1 \right\}. \quad (3.12)$$

We remark that using a different sampling rate is only for the purpose of simplifying the analysis for the special case where $k = 1$. In the actual algorithm, we will stick to the sampling rate described in Chapter 3.2.

We have the following theorem for the single-site case, that we will use to prove the results for the multi-site case.

Theorem 3.3.4. *For $k = 1$ and for a sufficiently large constant $\alpha > 0$, the randomized algorithm Non-monotonic Counter with the sampling probability as in (3.12) guarantees to continually track the count within relative accuracy ϵ with at least probability $1 - O(1/n)$ and uses the total expected communication of $O(\min\{1/(\mu\epsilon) \cdot \log^{3/2} n, \sqrt{n}/\epsilon \cdot \log n, n\})$ messages.*

Proof. In some parts of the analysis, we assume that $p \geq 1/2$, i.e. the drift μ is non-negative, which is sufficient as the other case can be analyzed by similar steps. The proof will follow the same main steps as in the proof of Theorem 3.3.1 and we will reuse the notation introduced therein.

Communication cost. We shall upper bound $\sum_{t \leq n} \mathbb{E}[R_t]$ which is the expected communication cost of our algorithm. Let us introduce two Bernoulli random vari-

ables R_t^1 and R_t^2 with respective parameters $\min\{\frac{\alpha \log^2 n}{\epsilon^2 S_t^2}, 1\}$ and $\min\{\frac{\alpha \log^3 n}{\epsilon t}, 1\}$. We then have

$$\sum_{t \leq n} \mathbb{E}[R_t] \leq \sum_{t \leq n} \mathbb{E}[R_t^1] + \sum_{t \leq n} \mathbb{E}[R_t^2].$$

Since

$$\sum_{t \leq n} \mathbb{E}[R_t^2] = O\left(\frac{\log n}{\epsilon}\right) \quad (3.13)$$

we shall focus on the term $\sum_{t \leq n} \mathbb{E}[R_t^1]$. Using the trick presented in Theorem 3.3.1, we have

$$\mathbb{E}[R_t^1] = \Pr[|S_t| \leq \vartheta] + \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right]$$

where $\vartheta = \frac{\sqrt{\alpha}}{\epsilon} \cdot \log n$.

We first note that $\mathbb{E}[R_t^1] = O(\sqrt{n}/\epsilon \cdot \log n)$ for any drift $\mu \in (-1, 1)$. The key observation is the simple fact that $\Pr[S_t = s] = O(1/\sqrt{t})$ for every $s \in [-t, t]$ which, for example, follows from the Berry-Esseen central limit theorem. Now, it easily follows $\sum_{t \leq n} \Pr[|S_t| \leq \vartheta] = \sum_{t \leq n} O(\vartheta/\sqrt{t}) = O(\vartheta\sqrt{n}) = O(\sqrt{n}/\epsilon \cdot \log n)$. Also, we have $\sum_{t \leq n} \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right] = O(\vartheta^2 \sum_{t \leq n} \sum_{s > \vartheta} \frac{1}{\sqrt{ts^2}}) = O(\vartheta\sqrt{n}) = O(\sqrt{n}/\epsilon \cdot \log n)$. We have thus showed that for every drift $\mu \in (-1, 1)$, we have

$$\sum_{t \leq n} \mathbb{E}[R_t^1] = O(\sqrt{n}/\epsilon \cdot \log n). \quad (3.14)$$

We next show that

$$\sum_{t \leq n} \Pr[|S_t| \leq \vartheta] = O\left(\frac{\vartheta}{\mu}\right). \quad (3.15)$$

Let us denote $\sigma^2 = 1 - \mu^2$ and $\rho = \sqrt{\frac{1+\mu}{1-\mu}}$. We observe

$$\Pr[|S_t| \leq \vartheta] = \sum_{s=-\vartheta}^{\vartheta} \binom{t}{\frac{t+s}{2}} 2^{-t} \sigma^t \rho^s = O\left(\frac{1}{\sqrt{t}} \cdot \sigma^t \sum_{s=-\vartheta}^{\vartheta} \rho^s\right) = O\left(\frac{1}{\sqrt{t}} \cdot \sigma^t \cdot \frac{\rho^{\vartheta+1} - 1}{\rho - 1}\right)$$

where we made use of the fact $\left(\frac{t}{2}\right)2^{-t} \leq \left(\frac{t}{2}\right)2^{-t} \leq \frac{\sqrt{2}}{\sqrt{\pi}\sqrt{t}}$. Now,

$$\frac{\rho^{\vartheta+1} - 1}{\rho - 1} = \frac{\mu^{\vartheta}}{\mu}(1 + o(1)) = \vartheta(1 + o(1)).$$

Therefore, $\Pr[|S_t| \leq \vartheta] = O(\vartheta\sigma^t/\sqrt{t})$. Observing that

$$\sum_{t=1}^n \frac{1}{\sqrt{t}}\sigma^t \leq e^{-\log(\frac{1}{\sigma})} + \int_1^n \frac{1}{\sqrt{t}}e^{-\log(\frac{1}{\sigma})t}dt = \Theta\left(\frac{1}{\log^{1/2}(\frac{1}{\sigma^2})}\right) = \Theta\left(\frac{1}{\mu}\right)$$

we obtain $\sum_{t \leq n} \Pr[|S_t| \leq \vartheta] = O(\vartheta/\mu)$, thus (3.15) holds.

We will further show that

$$\sum_{t \leq n} \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right] = O\left(\frac{\vartheta}{\mu} \cdot \log^{3/2} n\right). \quad (3.16)$$

We will separately consider two possible cases $\mu = o(\epsilon/\log n)$ and $\mu = \Omega(\epsilon/\log n)$.

Case 1: $\mu = o(\epsilon/\log n)$, i.e. $\mu = o(1/\vartheta)$. In this case, let $\tau_n = 1/\mu^2 \cdot \log^\delta n$, for $\delta > 1$.

Notice that we have

$$\begin{aligned} \sum_{t \leq \tau_n} \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right] &= \sum_{t \leq \tau_n} \vartheta^2 \sum_{s > \vartheta} \frac{1}{s^2} \Pr[S_t = s] \\ &= \sum_{t \leq \tau_n} \vartheta^2 \sum_{s > \vartheta} \frac{1}{s^2} \cdot O(1/\sqrt{t}) \\ &= O(\vartheta\sqrt{\tau_n}) = O\left(\frac{\vartheta}{\mu} \cdot \log^{\delta/2} n\right). \end{aligned}$$

Therefore, it remains to upper bound $\sum_{t > \tau_n} \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right]$. To do this, we observe that for every time $t \geq \tau_n$, it holds

$$\vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| > \vartheta)\right] \leq \Pr[S_t \leq \mu t/2] + \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(S_t > \mu t/2)\right].$$

For the first term, we use the Hoeffding bound to obtain

$$\Pr[S_t \leq \mu t/2] \leq \exp\left(-\frac{(\mu t/2)^2}{2t}\right) \leq \exp\left(-\frac{\mu^2 \tau_n}{8}\right) = \exp\left(-\frac{1}{8} \log^\delta n\right)$$

and observe that $\Pr[S_t \leq \mu t/2] = O(1/n^c)$, for arbitrary constant $c > 0$, and hence $\Pr[S_t \leq \mu t/2] = O(\vartheta/\mu)$.

It remains only to show $\sum_{t \geq \tau_n} \vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(S_t > \mu t/2)] = O(\vartheta/\mu)$, which we do as follows

$$\begin{aligned} \sum_{t > \tau_n} \vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(S_t > \mu t/2)] &= \sum_{t > \tau_n} \vartheta^2 \sum_{s > \mu t/2} O(\frac{1}{s^2 \sqrt{t}}) \\ &= \sum_{t > \tau_n} \vartheta^2 \cdot O(\frac{1}{\mu t^{3/2}}) = O(\frac{\vartheta^2}{\mu \tau_n^{1/2}}) \\ &= O(\frac{\vartheta^2}{\log^{\delta/2} n}) = o(\frac{\vartheta}{\mu}) \end{aligned}$$

where the last equality follows because of our assumption $\mu = o(1/\vartheta)$.

Case 2: $\mu = \Omega(\epsilon/\log n)$, i.e. $\mu = \Omega(1/\vartheta)$. Since we assume that the drift is positive, we have

$$\sum_{t \leq n} \vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(|S_t| > \vartheta)] \leq \sum_{t \leq n} 2\vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(S_t > \vartheta)].$$

Let us define $\tau_n = 2\vartheta/\mu$ and then note

$$\sum_{t \leq n} \vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(|S_t| > \vartheta)] \leq O(\frac{\vartheta}{\mu}) + \sum_{t \geq \tau_n} 2\vartheta^2 \mathbb{E}[\frac{1}{S_t^2} I(S_t > \vartheta)].$$

In the prevailing case $\mu = \Omega(1/\vartheta)$, so let $\mu \geq c/\vartheta$, for a constant $c > 0$ and sufficiently large n . We then have

$$\begin{aligned} \mathbb{E}[\frac{1}{S_t^2} I(S_t > \vartheta)] &\leq \mathbb{E}[\frac{1}{S_t^2} I(S_t > c/\mu)] \\ &= \frac{1}{\mu^2 t^2} \left(\sum_{c/\mu < s \leq \mu t/2} \frac{\mu^2 t^2}{s^2} \Pr[S_t = s] + 4 \Pr[S_t > \mu t] \right) \\ &\leq \frac{1}{\mu^2 t^2} \left(\frac{(\mu^2 t)^2}{c^2} \Pr[S_t \leq \mu t/2] + O(1) \right). \end{aligned}$$

Now, using the Hoeffding's inequality, we have $\Pr[S_t \leq \mu t/2] = \Pr[-S_t + \mu t \geq \mu t/2] \leq \exp(-\frac{\mu^2 t}{8})$, and thus it follows

$$\begin{aligned} \mathbb{E}[\frac{1}{S_t^2} I(S_t > \vartheta)] &\leq \frac{1}{\mu^2 t^2} \left(\frac{(\mu^2 t)^2}{c^2} e^{-\frac{(\mu^2 t)}{8}} + O(1) \right) \\ &= \frac{1}{\mu^2 t^2} \cdot O(1) = O\left(\frac{1}{\mu^2 t^2}\right). \end{aligned}$$

Therefore,

$$\sum_{t \geq \tau_n} \vartheta \mathbb{E}[\frac{1}{S_t^2} I(S_t > \vartheta)] = O(\vartheta^2 \sum_{t \geq 2\vartheta/\mu} \frac{1}{\mu^2 t^2}) = O\left(\frac{\vartheta}{\mu}\right).$$

Combining (3.13), (3.14), (3.15) and (3.16) with the trivial fact $\sum_{t \leq n} \mathbb{E}[R_t] \leq n$, we showed that the communication complexity is $\sum_{t \leq n} \mathbb{E}[R_t] = O(\min\{\frac{\log^{3/2} n}{\mu \epsilon}, \frac{\sqrt{n} \cdot \log n}{\epsilon}, n\})$.

Correctness. The probability that an error occurs is upper bounded as follows:

$$\Pr[E_n > 0] \leq \sum_{t \leq n} \mathbb{E}[R_t \cdot I(V_t < U_t)].$$

We shall show the following holds

$$\mathbb{E}[R_t \cdot I(V_t < U_t)] = O(1/n^2), \text{ for every } 1 \leq t \leq n \quad (3.17)$$

and then the proof of correctness follows via the union bound.

As already showed in (3.3), we have that for any value $r \geq 0$, the following holds

$$\Pr[V_t < U_t, V_t > r \mid S_t = s] \leq \Pr[V_t \leq r \mid S_t = s] + \Pr[r < U_t \mid S_t = s, V_t > r]. \quad (3.18)$$

Recall that $U_t \leq r$ means that at least one report is sent to the coordinator and $V_t > r$ means that the walk does not escape the ball $\mathbf{B}_\epsilon(s)$ within next r updates as observed at update t . Similar to our analysis in Theorem 3.3.1, the “trick” here is to identify an appropriate value of r (that could depend on s) such that w.h.p. between update t and $t + r$, the following two events happen:

1. at least one message is sent, and
2. the walk has not escape the ball $\mathbf{B}_\epsilon(s)$.

Similar to as in in equations (3.4) and (3.5) in the proof of Theorem 3.3.1, we will next choose the value of r such that both above conditions hold with high probability. Specifically, we take

$$r = \min \left\{ \frac{\epsilon s}{(1 - \epsilon)^2}, \frac{1}{\mu} \right\} \cdot \frac{\epsilon s}{c \log n}$$

where c is a large enough positive constant. Notice again that the variable r is chosen only for the purpose of analysis and is not used for the execution of the algorithm.

In the following we first upper bound the probability that the count exits the ball $\mathbf{B}_\epsilon(s)$ and then upper bound the probability that no message is sent to the coordinator.

Escape from the ball $\mathbf{B}_\epsilon(s)$. We may decompose the subwalk into two components: the drift component and the noise component. We shall show that for every $r' \leq r$, the following holds:

1. The expected increment is smaller than $\epsilon s/2$, i.e.

$$\mathbb{E} \left[\sum_{t \leq i \leq t+r'} X_i \right] \leq \epsilon s/2. \quad (3.19)$$

2. The value of the increment is within $\mu r' \pm \epsilon s/2$ with high probability, i.e.

$$\Pr \left[\left| \sum_{t \leq i \leq t+r'} X_i - \mu r' \right| \geq \epsilon s/2 \right] \text{ is small.} \quad (3.20)$$

When both two conditions hold, we are able to show that the probability that the walk escapes the ball $\mathbf{B}_\epsilon(s)$ is small.

Condition 1: Observe that $r \leq \epsilon s / (\mu c \log n)$. Therefore, the expected drift between time t and $t + r'$ is at most

$$\mu r \leq \epsilon s / (c \log n) \leq \epsilon s / 2.$$

Condition 2: Let us denote the noise component by $N \triangleq \sum_{t \leq i \leq t+r'} X_i - r' \mu$. By using the Hoeffding inequality, we have for large enough constant $c > 0$,

$$\Pr[|N| \geq \epsilon s / 2] \leq 2 \exp\left(-\frac{(\epsilon s)^2}{8r'}\right) \leq 2 \exp\left(-\frac{(\epsilon s)^2}{8r}\right) = O(1/n^3).$$

By using the union bound, we can conclude that the probability that the walk escapes $\mathbf{B}_\epsilon(s)$ is $O(1/n^2)$, i.e. $\Pr[V_t \leq r \mid S_t = s] = O(1/n^2)$.

Time until next message. We now upper bound the probability $\Pr[r < U_t \mid S_t = s, V_t > r]$ by considering two possible cases.

Case 1. $\frac{\epsilon s}{(1-\epsilon)^2} \leq \frac{1}{\mu}$, i.e. $r = \frac{(\epsilon s)^2}{(1-\epsilon)^2 c \log n}$. In this case, the noise component is dominant. The sampling rate is at least as large as $p_{t'} \geq \min\left\{\frac{\alpha \log^2 n}{(1-\epsilon)^2 (\epsilon s)^2}, 1\right\}$ for every $t < t' \leq t + r$. If $p_{t'} = 1$ for at least one $t < t' \leq t + r$, we have $\Pr[r < U_t \mid S_t = s, V_t > r] = 0$ and we are done. If, otherwise, for every $t < t' \leq t + r$ it holds $p_{t'} < 1$, then we have

$$\Pr[r < U_t \mid S_t = s, V_t > r] \leq \left(1 - \frac{\alpha \log^2 n}{(1-\epsilon) \epsilon^2 s^2}\right)^r. \quad (3.21)$$

By substituting the value $r = \frac{(\epsilon s)^2}{(1-\epsilon)^2 c \log n}$, we have $\Pr[r < U_t \mid S_t = s, V_t > r] = O(1/n^2)$, for sufficiently large constant c .

Case 2. $\frac{\epsilon s}{(1-\epsilon)^2} > \frac{1}{\mu}$, i.e. $r = \frac{\epsilon s}{\mu c \log n}$. In this case, the drift term is dominant. The sampling rate satisfies $p_{t'} \geq \min\left\{\frac{c \log^3 n}{\epsilon t}, 1\right\}$ for every $t < t' \leq t + r$. In case that

there exists t' such that $p_{t'} = 1$, we have $\Pr[r < U_t \mid S_t = s, V_t > r] = 0$ and we are done. If for every $t < t' \leq t + r$, $p_{t'} < 1$, we have

$$\begin{aligned} \Pr[r < U_t \mid S_t = s, V_t > r] &\leq \left(1 - \frac{\alpha \log^3 n}{\epsilon t}\right) \left(1 - \frac{\alpha \log^3 n}{\epsilon(t+1)}\right) \cdots \left(1 - \frac{\alpha \log^3 n}{\epsilon(t+r)}\right) \\ &\leq \left(1 - \frac{\alpha \log^3 n}{\epsilon(t+r)}\right)^r. \end{aligned}$$

We further consider two subcases: when t is small and otherwise.

Case 2a. $t < \frac{\alpha r \log^2 n}{\epsilon}$. For this case, we have $t + r \leq 2\alpha r \log^2 n / \epsilon$, and thus

$$\left(1 - \frac{\alpha \log^3 n}{(t+r)\epsilon}\right)^r \leq \left(1 - \frac{\alpha \log^3 n}{2\alpha r \log^2 n}\right)^r = O(1/n^2).$$

Case 2b. $t > \frac{\alpha r \log^2 n}{\epsilon}$. For this case $t + r < 2t$. The crucial observation is that the value of S_t has a strong lower bound, i.e. it holds with high probability. Specifically, by the Hoeffding inequality, we have

$$\Pr[S_t \leq \frac{t\mu}{2}] \leq \exp\left(-\frac{\mu^2 t}{8}\right).$$

Using the facts $t > (\alpha r \log^2 n)/\epsilon$, $r = \frac{\epsilon s}{c\mu \log n}$ and $s > \frac{(1-\epsilon)^2}{\mu\epsilon}$, we have $t \geq \frac{\alpha}{c\mu^2}(1 - \epsilon)^2 \log n$, i.e.

$$\Pr[S_t \leq \frac{t\mu}{2}] = O(1/n^2). \quad (3.22)$$

Therefore, we have for $s \geq \frac{t\mu}{2}$,

$$\Pr\left[r < U_t \mid S_t = s, V_t > r\right] \leq \left(1 - \frac{c \log^3 n}{2t\epsilon}\right)^{\epsilon t / (c \log n)} \leq \exp\left(-\frac{\alpha}{c} \cdot \log^2 n\right) \leq 1/n^2.$$

Summary. Finally, we summarize the above cases. Let e be the following event

$$e = \left\{S_t > \frac{(1-\epsilon)^2}{\mu\epsilon}\right\} \cap \left\{S_t < \frac{t\mu}{2}\right\}.$$

Equation (3.22) gives us $\Pr[\neg e] = O(1/n^2)$. Also, let f be an element in the probability space associated with \mathcal{F}_t and $F_{\mathcal{F}_t}(\cdot)$ be the distribution function of \mathcal{F}_t . We note that the following series of inequalities hold

$$\begin{aligned}
& \mathbb{E}_{\mathcal{F}_t} [R_t \cdot \Pr[V_t < U_t \mid S_t]] \\
&= \int_f R_t \Pr[V_t < U_t \mid f] dF_{\mathcal{F}_t}(f) \\
&= \int_f R_t \Pr[V_t < U_t \mid f] dF_{\mathcal{F}_t}(f, e) + \int_f R_t \Pr[V_t < U_t \mid f] dF_{\mathcal{F}_t}(f, \neg e) \\
&\quad \text{(rule out the rare event } \neg e.) \\
&\leq \int_f R_t \Pr[V_t < U_t \mid f] dF_{\mathcal{F}_t}(f, e) + O(1/n^2) \\
&\leq \int_f R_t (\Pr[V_t < r \mid f] + \Pr[r \leq U_t \mid f, V_t \geq r]) dF_{\mathcal{F}_t}(f, e) + O(1/n^2) \\
&\leq \int_f R_t (O(1/n^2) + O(1/n^2)) dF_{\mathcal{F}_t}(f, e) + O(1/n^2) \\
&\leq O(1/n^2) \int_f R_t dF_{\mathcal{F}_t}(f, e) + O(1/n^2) \\
&\leq O(1/n^2) \cdot \mathbb{E}[R_t] = O(1/n^2).
\end{aligned}$$

Therefore, $\sum_{t \leq n} \mathbb{E}[R_t \cdot I(V_t < U_t)] = O(1/n)$ which completes the proof. \square

Multiple site case. We now move to the analysis for the multiple site case. Similar to the strategy we used in Chapter 3.3.1, we split the no-drift phase into the straightforward (StraightSync) stage and the broadcast (SBC) stage, as discussed in the previous section. We then have the following theorem.

Theorem 3.3.5. *There exists a choice of constants α and $\beta > 0$ for the randomized algorithm Non-monotonic Counter, for the k -site count problem with unknown drift, to guarantee the continuous tracking within a prescribed relative accuracy ϵ with high probability and the following communication cost in expectation:*

- $\tilde{O}\left(\min\left\{\frac{\sqrt{k}}{|\mu|\epsilon}, \frac{\sqrt{kn}}{\epsilon}, n\right\} + \frac{\sqrt{k}}{\epsilon}\right)$, if $k = O(1/(\mu\epsilon)^2)$, and
- $\tilde{O}\left(\min\left\{\frac{\sqrt{k}}{|\mu|\epsilon}, \frac{\sqrt{kn}}{\epsilon}, n\right\} + k\right)$, if $k = \omega(1/(\mu\epsilon)^2)$.

Notice that our algorithm's communication cost has two types of asymptotic behaviors for different k because the HYZ counter uses different strategies for different k .

Proof of Theorem 3.3.5. We shall show the following three statements in our analysis:

1. Using the non-drift algorithm before update $\tau = \Theta(1/(\mu^2\epsilon) \cdot \log n)$ is sufficient to track the count with high probability.
2. Using the difference estimator after update τ is also sufficient to track the count with high probability.
3. The communication complexity in the first phase is $\tilde{O}(\min\{\frac{\sqrt{k}}{|\mu|\epsilon}, \frac{\sqrt{kn}}{\epsilon}\})$.

Correctness up to time τ . To show the first part, since our k -site algorithm is only mimicking the algorithm for the 1-site case, we only need to show that our sampling rate here is at least as large as the sampling rate for the 1-site without drifting case. Specifically, we shall show that

$$\min\left\{\frac{\alpha \log^4 n}{\epsilon^2 S_t^2}, 1\right\} \geq \min\left\{\frac{\alpha \log^2 n}{\epsilon^2 S_t^2} + \frac{\alpha \log^3 n}{\epsilon t}, 1\right\}$$

with high probability, i.e., $\log^4 n/(\epsilon^2 S_t^2) = \Omega(\log^3 n/(\epsilon t))$, and hence, $S_t = O(\sqrt{t} \log n/\epsilon)$ with high probability. This can be seen through applying a standard Chernoff bound.

Correctness after time τ . Let P_t and M_t denote the number of positive and negative increments observed up to time t , respectively, i.e. $P_t = \sum_{i \leq t} I(X_i = 1)$ and $M_t = \sum_{i \leq t} I(X_i = -1)$. Let \hat{P}_t and \hat{M}_t be estimators of P_t and M_t , respectively, such that $\hat{P}_t \in \mathbf{B}_{1/(\epsilon\mu)}(P_t)$ and $\hat{M}_t \in \mathbf{B}_{1/(\epsilon\mu)}(M_t)$. Here, we show that $\hat{P}_t - \hat{M}_t \geq (1 - \epsilon)S_t$ with high probability. Notice that the other condition $\hat{P}_t - \hat{M}_t \leq (1 + \epsilon)S_t$ can also be showed to hold with high probability in a similar fashion.

First, notice that $\hat{P}_t - \hat{M}_t \geq P_t - M_t - \mu\epsilon t/6$. By using a Chernoff bound and the fact that $t = \Omega(1/(\mu^2\epsilon) \cdot \log n)$, we have $t \leq \frac{2}{3\mu}S_t$ with high probability. Therefore, $\hat{P}_t - \hat{M}_t \geq P_t - M_t - \epsilon S_t$ with high probability.

Communication cost for the first phase. We reuse the notation developed for the analysis of multiple site case with no drift. Let C_t be the communication cost associated with the t -th update. We have

$$\mathbb{E}[C_t] \leq \Pr[S_t \leq \sqrt{k}/\epsilon] + \frac{k \log^4 n}{\epsilon^2} \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| \leq \sqrt{k}/\epsilon)\right]. \quad (3.23)$$

Using the communication analysis developed in Theorem 3.3.4 and re-scaling the parameter ϵ as ϵ/\sqrt{k} , we have

$$\begin{aligned} & \Pr[|S_t| \leq \vartheta] + \vartheta^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| \geq \vartheta)\right] \\ &= \Pr[S_t \leq \sqrt{2\alpha}\sqrt{k}/\epsilon \log^2 n] + \frac{k \log^2 n}{\epsilon^2} \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| \geq \sqrt{k}/\epsilon)\right] \\ &\geq \frac{1}{\log^2 n} \mathbb{E}[C_t] \end{aligned}$$

where the last inequality follows from (3.23).

Therefore,

$$\sum_{t \leq n} \mathbb{E}[C_t] \leq \log^2 n \left(\sum_{t \leq n} (\Pr[|S_t| \leq \vartheta_n] + \vartheta_n^2 \mathbb{E}\left[\frac{1}{S_t^2} I(|S_t| \geq \vartheta_n)\right]) \right)$$

which allows us to conclude the communication cost at the first phase is indeed $\tilde{O}(\min\{\frac{\sqrt{k}}{|\mu|\epsilon}, \frac{\sqrt{kn}}{\epsilon}\})$.

□

3.3.3 Randomly Ordered Data Streams

We now move to the random permutation case. We use the same tracking algorithm described in Chapter 3.2 to solve this problem by using the sampling rate defined in (3.1) with $\beta = 2$ and sufficiently large $\alpha > 0$.

Theorem 3.3.6. *Let a_1, \dots, a_n be an arbitrary, randomly permuted, sequence of bounded real values. The randomized algorithm Non-monotonic Counter with the sampling probability in (3.1) for $\beta = 2$ and sufficiently large constant $\alpha > 0$ guarantees to track the count within the relative accuracy ϵ with probability $1 - O(1/n)$ and uses the total expected communication of $O(\sqrt{kn}/\epsilon \cdot \log n + \log^3 n)$ messages.*

Note that here, because of the adversarial choice of the input sequence, we cannot exploit the drift. We remark that when the update is a fractional number from $[-1, 1]$ rather than $\{-1, 1\}$, our Non-monotonic Counter algorithm still holds. The key difference between the analysis for Theorem 3.3.6 and the one for i.i.d. input is that the updates are correlated when the content of the stream is decided in advance. This difference boils down to a modified analysis for the first passage time of the partial sums. In the Bernoulli i.i.d. case, a straightforward application of Hoeffding's inequality suffices to give a bound on the first passage time. While here Hoeffding's inequality is no longer applicable, we are able to use tail inequalities for sampling without replacement [56, 98] to circumvent this problem.

In what follows, we first analyze the algorithm for the single site case. Then we shall proceed to prove Theorem 3.3.6.

Single site case.

Let us recall that μ is used to represent the drift rate of the counter process, which for a random permutation is defined by letting μn denote the final value of the counter.

Our algorithm for the single site case is identical to the single site algorithm for i.i.d. input stream. Specifically, we shall use the following sampling rate, for a constant $\alpha > 0$,

$$\text{Sample-Prob}(S_t, t) = \min \left\{ \frac{\alpha \log^2 n}{\epsilon^2 S_t^2} + \frac{\alpha \log^3 n}{\epsilon t}, 1 \right\}. \quad (3.24)$$

We next prove Theorem 3.3.6 for the special case $k = 1$, which is decomposed into three steps. In Lemma 3.3.9, we show that the algorithm is correct for sufficiently small drift. In Lemma 3.3.10, we show that the algorithm is correct for large drift. Finally, we shall analyze the communication cost in Lemma 3.3.11.

We first present two lemmas about deviation of a random permutation from a typical sample-path which we shall use in the remainder of the proof.

Lemma 3.3.7. *Let X_1, X_2, \dots, X_n be a random permutation such that $\sum_{i \leq n} X_i = \mu n$. Also, let $S_t \triangleq \sum_{i=1}^t X_i$. Then, we have for every $c \geq 0$,*

$$\Pr[\exists t < \frac{n}{2} : |S_t - \frac{t}{n} \mu| \geq \sqrt{2(1+c)t \log n}] \leq \frac{1}{n^c}. \quad (3.25)$$

We may apply Lemma 3.3.7 to the left half and right half of permutation (when we apply it to the right half, we read the permutation from right to left) and yield the following corollary.

Corollary 3.3.8. *Let X_1, X_2, \dots, X_n be a random permutation such that $\sum_{i=1}^n X_i = \mu n$. Also, let $S_t \triangleq \sum_{i=1}^t X_i$. Then, we have for every $c \geq 0$,*

$$\Pr[\exists t < \frac{n}{2} : |S_t - \frac{t}{n}\mu| \geq \sqrt{2(1+c) \min\{t, n-t\} \log n}] \leq \frac{1}{n^c}. \quad (3.26)$$

Proof. (of Lemma 3.3.7) Our proof is a straightforward application of Hoeffding inequality for the special case of sampling *without* replacement. Specifically, by Theorem 2 in [56], we have for every $t \leq \frac{n}{2}$ and $c \geq 0$,

$$\Pr[|S_t - \frac{t}{n}\mu| \geq \sqrt{2(1+c)t \log n}] \leq 2 \exp(-\frac{(\sqrt{2(1+c)t \log n})^2}{2t}) = \frac{2}{n^{1+c}}.$$

By using the union bound across t under consideration, we complete the proof. \square

We next state our main result for the case where $\mu = \tilde{O}(\sqrt{n})$.

Lemma 3.3.9. *Let μn be the final value of the counter. Suppose that $\mu n \leq \sqrt{n} \log^2 n$, then the algorithm Non-monotonic Counter with the sampling probability as in (3.24) correctly tracks the counter with probability $1 - O(1/n)$.*

Proof. We again use the same notation as for the i.i.d. case. We have

$$\Pr[E_n > 0] \leq \sum_{t \leq n} \mathbb{E}[R_t \cdot I(V_t < U_t)].$$

Our goal is again to show that $\mathbb{E}[R_t \cdot I(V_t < U_t)] = O(1/n^2)$. We need to consider two cases $t \leq n/2$ and $t > n/2$. In what follows, we only focus on the case $t \leq n/2$ as the case $t > n/2$ can be analyzed by following similar same steps.

Define the event e that for large enough constant $\gamma > 0$, the following holds: $|S_t - \mu t| \leq \gamma \sqrt{t \log n}$ for every $1 \leq t \leq n$. Lemma 3.3.7 gives us $\Pr[e] \geq 1 - 1/n^{\frac{1}{2}\gamma^2 - 1}$. When the event e occurs, we have $S_t \geq -\gamma \sqrt{t \log n}$.

In what follows, our analysis will focus on the case where e happens. The probability of $\neg e$ is sufficiently small that it becomes negligible for our analysis.

Recall that

$$\mathbb{E}[R_t I(V_t < U_t)] = \mathbb{E}_{\mathcal{F}_t}[R_t \Pr[V_t < U_t \mid S_t]].$$

and that for any $r \geq 0$, the following holds

$$\Pr[V_t < U_t \mid S_t = s] \leq \Pr[V_t \leq r \mid S_t = s] + \Pr[r < U_t \mid S_t = s, V_t > r]. \quad (3.27)$$

Let us define $r = \frac{\epsilon^2 s^2}{c(1-\epsilon)^2 \log n}$, where $c > 0$ is an appropriately chosen constant. We need to analyze two terms that appear in the right-hand side of equation (3.27) which, respectively, correspond to the probability that the walk between steps t and $t + r$ exits the ball $\mathbf{B}_\epsilon(s)$ and the probability that no message is sent between t and $t + r$. Our goal is to show that both terms are $O(1/n^2)$, which allows us to conclude the correctness proof.

Escape from the ball $\mathbf{B}_\epsilon(s)$ Here, we need to bound $\Pr[V_t \leq r \mid S_t = s, e]$. We shall again view the walk as a composition of a drift and noise component and show that the values of both components are smaller than $\epsilon s/2$ from update t to update $t + r$ in order to conclude that $\Pr[V_t \leq r \mid S_t = s, e]$ is small.

Case 1: drifting is small. Now, let us consider the trajectory of the walk from t to n . We know that the walk starts at value s and ends at value μn . Since no information during this period is revealed, the trajectory is a random trajectory (with known starting and ending point). In other words, we may view the trajectory as a random permutation with drift $\mu n - s$. Since $s \geq -\gamma\sqrt{t \log n}$ (recall that we

assume event e happens), we have that the total drift for the trajectory is bounded by $\mu n + \gamma\sqrt{n \log n} \leq 2\sqrt{n} \log^2 n$. The “drift rate” thus is $\leq \frac{4}{\sqrt{n}} \log^2 n$. Next, we compute an upper bound for the expected increment until time $t + r'$, for $0 < r' \leq r$. We have, for every $0 < r' \leq r$,

$$\mathbb{E}\left[\sum_{t \leq i \leq t+r'} X_i\right] \leq \frac{4}{\sqrt{n}} \log^2 n \cdot r'.$$

Case 2: noise is small. Let us define $N = \mathbb{E}[\sum_{t \leq i \leq t+r'} X_i] - \mu r'$. By using the Hoeffding inequality (without replacement), we have

$$\Pr[|N| \geq \frac{\epsilon s}{2}] \leq 2 \exp\left(-\frac{\epsilon^2 s^2}{8r'}\right) \leq 2 \exp\left(-\frac{\epsilon^2 s^2}{8r}\right) = O(1/n^3).$$

By using the union bound, we may conclude that

$$\Pr\left[\left|\sum_{t \leq i \leq t+r'} X_i\right| > \epsilon s, \text{ for some } 1 \leq r' \leq r\right] = O(1/n^2).$$

Time until next message. We now move to the term $\Pr[V_t < U_t, V_t > r \mid S_t = s, e]$. By using the same manipulation as in the proof of Theorem 3.3.3, we have that for large enough constant $c > 0$,

$$\Pr[V_t < U_t, V_t > r \mid S_t = s, e] \leq \Pr[r \leq U_t \mid S, V_t > r] \leq \left(1 - \frac{(1-\epsilon)^2 c \log^2 n}{\epsilon^2 s^2}\right)^r = O(1/n^2).$$

Therefore, we may conclude that

$$\Pr[R_t \cdot I(V_t < U_t)] \leq \Pr[\neg e] + \int_f \mathbb{E}[I(V_t < U_t) \mid f] dF_{\mathcal{F}_t}(f, e) = O(1/n^2)$$

which completes our analysis. \square

Now we proceed to analyze the case where $\mu > \sqrt{n} \log^2 n$.

Lemma 3.3.10. *Let μ be the final value of the counter. Suppose $\mu n \geq \sqrt{n} \log^2 n$, then the algorithm Non-monotonic Counter with the sampling probability as in (3.24) correctly tracks the counter with probability $1 - O(1/n)$.*

Proof. We shall use the same set of notation and start by recalling the fact

$$\Pr[E_n > 0] \leq \sum_{t \leq n} \mathbb{E}[R_t I(V_t < U_t)].$$

We need to show that $\mathbb{E}[R_t I(V_t < U_t)] = O(1/n^2)$ to conclude our analysis. Recall that we use e to represent the event that $|S_t - \mu t| \leq \gamma\sqrt{t \log n}$, for every $1 \leq t \leq n$, where γ is a large enough constant. Using

$$\Pr[V_t < U_t \mid S_t = s, e] \leq \Pr[r < U_t \mid S_t = s, V_t > r, e] + \Pr[V_t \leq r \mid S_t = s, e],$$

we need to show that $\Pr[r < U_t \mid S_t = s, V_t > r, e] = O(1/n^2)$ and $\Pr[V_t \leq r \mid S_t = s, e] = O(1/n^2)$. In this case, we choose r as follows

$$r = \min \left\{ \frac{(\epsilon s)^2}{c(1 - \epsilon)^2 \log n}, \frac{\epsilon s}{c\mu \log n} \right\}$$

where c is a large enough constant $c > 0$.

Escape from the ball $\mathbf{B}_\epsilon(s)$ We now analyze the term $\Pr[V_t \leq r \mid S_t = s, e]$. We again view the random walk as a composition of the drifting and noise component and show that the values of both components are smaller than $\epsilon s/2$ from update t to update $t + r$ in order to conclude that $\Pr[V_t \leq r \mid S_t = s, e]$ is small.

Case 1: drifting is small. Let $r' \leq r$. We shall again interpret the trajectory from t to n as a random permutation that starts at value s and ends at value μn . Since $s \geq \mu t - \gamma\sqrt{t \log n}$ (under event e), we have that the total drift is at most $\mu(n - t) + \gamma\sqrt{t \log n}$. We can verify that $\mu(n - t) > \gamma\sqrt{t \log n}$ for sufficiently large n . Therefore, $2\mu(n - t)$ is an upper bound of the total drift in the random trajectory part. Now, for any $0 < r' \leq r$, the expected increment between t and $t + r'$ is

$\mathbb{E}[\sum_{t \leq i \leq t+r'} X_i] \leq 2\mu r \leq \epsilon s/2$, where the last inequality holds because $r \leq \frac{\epsilon s}{c\mu \log n}$.

Case 2: noise is small. For any $r' \leq r$, let us recall that $N = \mathbb{E}[\sum_{t \leq i \leq t+r'} X_i] - \mu r'$.

By using the Hoeffding inequality (without replacement), we have

$$\Pr[|N| \geq \frac{\epsilon s}{2}] \leq 2 \exp(-\frac{(\epsilon s)^2}{8r'}) \leq 2 \exp(-\frac{(\epsilon s)^2}{8r}) = O(1/n^3).$$

We may thus conclude that $\Pr[V_t \leq r \mid S_t = s, e] = O(1/n^2)$.

Time until next message. We now upper bound the probability $\Pr[r < U_t \mid S_t = t, V_t > r, e]$. We again separately consider two possible cases.

Case 1. $\frac{\epsilon s}{(1-\epsilon)^2} \leq \frac{1}{\mu}$, i.e. $r = \frac{(\epsilon s)^2}{c(1-\epsilon)^2 \log n}$. In this case, we have

$$\Pr[U_t > r \mid V_t > r, S_t = s] \leq \left(1 - \frac{\alpha \log^2 n}{(1-\epsilon)^2 (\epsilon s)^2}\right)^r = O(1/n^2), \text{ for large enough } \gamma > 0.$$

Case 2. $\frac{\epsilon s}{(1-\epsilon)^2} \geq \frac{1}{\mu}$, i.e. $r = \frac{\epsilon s}{c(\log n)\mu}$. In this case, we need to further consider two subcases: $t \leq 1/(\epsilon\mu^2) \cdot \log n$ or otherwise. Recall that $\tilde{O}(1/\mu^2\epsilon)$ is approximately the cut such that in the region before the cut, the noise term dominates and in the region after the cut the drifting term dominates.

Case 2a. $t \leq 1/(\epsilon\mu^2) \cdot \log n$. We have

$$\begin{aligned} \Pr[U_t > r \mid V_t > r, S_t = s] &\leq \left(1 - \frac{\gamma \log^3 n}{\epsilon t}\right)^r \\ &\leq \left(1 - \frac{\gamma \log^3 n}{\epsilon \log n / (\mu^2 \epsilon)}\right)^{\frac{\epsilon s}{c(1-\epsilon)^2 \log n \mu}} \\ &\leq \left(1 - \frac{\gamma \log^3 n}{\log n / (\mu^2)}\right)^{\frac{\epsilon \frac{1}{\mu \epsilon} (1-\epsilon)^2}{c(1-\epsilon)^2 \log n \mu}} \\ &= O(1/n^2) \end{aligned}$$

where in the second inequality we use the fact $t \leq 1/(\mu^2\epsilon) \cdot \log^2 n$ and the definition of r , and in the third inequality we use the fact $s \geq (1 - \epsilon)^2/(\mu\epsilon)$.

Case 2b. $t \geq 1/(\mu^2\epsilon) \cdot \log n$. The event e gives us $s \geq \mu t/2$, for sufficiently large n .

Therefore, $r \geq \frac{\epsilon\mu t/2}{c\mu \log n} = \frac{\epsilon t}{2c \log n}$. We then have

$$\Pr[r < U_t \mid S_t, V_t > r] \leq \left(1 - \frac{\gamma \log^3 n}{\epsilon t}\right)^{\frac{\epsilon t}{2c \log n}} = O(1/n^2).$$

□

We next move to analyze our communication cost. We have the following lemma.

Lemma 3.3.11. *The expected communication complexity of algorithm Non-monotonic Counter to track any random permutation is at most $O(\frac{1}{\epsilon} \cdot \sqrt{n} \log n + \log^3 n)$ if μ is bounded away from 1 (i.e. $1 - \mu = \Omega(1)$).*

Proof. We shall show that for any $t \in [\log^2 n, n - \log^2 n]$, it holds

$$\Pr[S_t = x] = O\left(\frac{1}{\min\{\sqrt{t}, \sqrt{n-t}\}}\right), \quad -t \leq x \leq t. \quad (3.28)$$

Then we may use the technique developed in Theorem 3.3.1 to conclude that the expected communication cost from the sampling component $\tilde{\Theta}(1/(\epsilon^2 s^2))$ is $O(\frac{1}{\epsilon} \cdot \sqrt{n} \log n)$ while that of the sampling component $\tilde{\Theta}(1/(\epsilon t))$ is $O(\log^3 n)$, which would complete our proof. Notice that giving up the regions $[0, \log^2 n]$ and $[n - \log^2 n, n]$ is legitimate because the communication cost incurred in these two regions is asymptotically dominated by that in the other region.

We now proceed to prove (3.28). Let x be the number of positive signs and y be the number of negative signs. We have $x + y = n$ and $x - y = \mu n$, i.e. $x = \frac{1+\mu}{2}n$ and $y = \frac{1-\mu}{2}n$. We divide the positive signs and negative signs into three groups, D ,

X , and Y . The set Y consists of the set of negative signs; the size of Y is $\frac{1-\mu}{2}n$. X consists of the set of positive signs of size $\frac{1-\mu}{2}\mu$. D consists of the remaining signs of size μn .

We use the following way of generating random permutations:

- We first place signs from D uniformly at random across all n time units.
- We next place X and Y uniformly at random in the rest of the slots.

Let D_t be the number of signs from D that fall into the region $[0, t]$. Then D_t follows a binomial distribution i.e. $D_t \sim B(\mu n, t/n)$. By using a Chernoff bound, we have for any constant ϵ_0 :

$$\Pr[D_t \geq (1 + \epsilon_0)\mu t] \leq \exp(-\epsilon_0^2\mu/3). \quad (3.29)$$

Via some straightforward algebra manipulation, we may conclude from (3.29) that

$$\Pr[t - D_t = \Omega(t)] \geq 1 - \frac{1}{n^4}.$$

The fraction $1/n^4$ is chosen rather arbitrary. Let us denote the event $t - D_t = \Omega(t)$ as e . Under event e , let $t_0 = t - D_t = \Theta(t)$. The rest of the slots in $[0, t]$ then are chosen uniformly from X and Y . Let x_0 be the number of signs that are chosen from X and y_0 be the number of signs that are chosen from Y , where $x_0 + y_0 = t_0$. We have

$$\Pr[S_t = D_t + x_0 - y_0] = \frac{\binom{|X|}{x_0} \binom{|Y|}{y_0}}{\binom{|X|+|Y|}{t_0}}.$$

Let us write $q(x_0) = \Pr[S_t = D_t + x_0 - y_0]$. We shall show that

1. $q(1) < q(2) < q(3) < \dots < q(t_0/2) > q(t_0/2 + 1) > \dots > q(t_0)$.
2. $q(\frac{t_0}{2}) = \Theta(q(\frac{t_0}{2} + 1)) = \Theta(q(\frac{t_0}{2} + 2)) = \dots = \Theta(q(\frac{t_0}{2} + \sqrt{t_0}))$.

The second item allows us to conclude that $q(t_0/2) = O(1/\sqrt{t_0}) = O(1/\sqrt{t})$. The first item gives us that $q(t_0/2)$ is the maximum item among all, which allows us to conclude that $\Pr[S_t = x] = O(1/\sqrt{t})$ for all x .

Proving the first item only require some straightforward manipulation. The second item seems to be a known folklore. For completeness, we sketch a proof for $q(t_0/2) = \Theta(q(t_0/2 + \sqrt{t_0}))$. Our goal is to show that

$$\frac{q(t_0/2)}{q(t_0/2 + \sqrt{t_0})} = O(1).$$

We have

$$\begin{aligned} & \frac{q(t_0/2)}{q(t_0/2 + \sqrt{t_0})} \\ = & \frac{\binom{|X|}{\frac{t_0}{2}} \binom{|Y|}{\frac{t_0}{2}}}{\binom{|X|}{\frac{t_0}{2} + \sqrt{t_0}} \binom{|Y|}{\frac{t_0}{2} - \sqrt{t_0}}} \\ = & \frac{(|X| \cdots |X - \frac{t_0}{2}|)^2}{[(t_0/2)!]^2} \cdot \frac{(t_0/2 - \sqrt{t_0})!(t_0/2 + \sqrt{t_0})!}{(|X| \cdots |X - \frac{t_0}{2} + \sqrt{t_0}|)(|X| \cdots |X - \frac{t_0}{2} - \sqrt{t_0}|)}. \end{aligned}$$

Notice that

$$\begin{aligned} \frac{(t_0/2 - \sqrt{t_0})!(t_0/2 + \sqrt{t_0})!}{[(t_0/2)!]^2} &= \frac{(t_0/2 + 1) \cdots (t_0/2 + \sqrt{t_0})}{(t_0/2 - \sqrt{t_0} + 1) \cdots t_0/2} \\ &= \prod_{i \leq \sqrt{t_0}} \left(1 + \frac{\sqrt{t_0}}{t_0/2 - \sqrt{t_0} + i} \right) \\ &\leq \left(1 + \frac{\sqrt{t_0}}{t_0/2} \right)^{\sqrt{t_0}} \\ &= O(1). \end{aligned}$$

Similarly, we may show that

$$\frac{(|X| \cdots |X - \frac{t_0}{2} + \sqrt{t_0}|)(|X| \cdots |X - \frac{t_0}{2} - \sqrt{t_0}|)}{(|X| \cdots |X - \frac{t_0}{2}|)^2} = O(1),$$

concluding $q(t_0/2) = O(q(t_0/2 - \sqrt{t_0}))$, which completes our complexity analysis. \square

Multiple site case. We now prove Theorem 3.3.6.

Proof of Theorem 3.3.6. Our analysis mostly follows the algorithm for the drifted case. Here we highlight a few key components in our analysis. Specifically, we need to verify the following statements.

1. It takes $\tilde{O}(1/\mu^2)$ samples to estimate μ within an arbitrarily fixed relative accuracy.
2. When $t = \tilde{O}(1/(\mu^2\epsilon))$, then $1/(\epsilon^2 S_t^2) = \tilde{\Omega}(1/\epsilon t)$. This condition makes sure that treating the input as non-drifting is fine before time $t = \tilde{O}(1/(\mu^2\epsilon))$.
3. When $t = \tilde{\Omega}(1/(\mu^2\epsilon))$, then the difference estimator works.
4. The expected communication cost is
 - (a) If $t = \tilde{O}(1/(\mu^2\epsilon))$, the expected communication cost is $\tilde{O}(\sqrt{\frac{k}{t}} \frac{1}{\epsilon})$.
 - (b) If $t = \tilde{\Omega}(1/(\mu^2\epsilon))$, the expected cost is
 - $\tilde{O}(\sqrt{k/(\mu\epsilon)})$ when $k < 1/(\mu^2\epsilon^2)$.
 - $\tilde{O}(k)$ otherwise.

First, notice that by Theorem A.1.5, when $t = \Omega(\log n/(\mu^2\epsilon))$, we have $\Pr[|S_t - \mu t| > 0.1\mu t] \leq \exp(-\Theta(t\mu^2)) = \exp(-\Theta(\log^2 n))$. Therefore, the concentration results still holds. Item 1 and 3 above can be proved by only using this tail inequality. For item 2, we essentially only need to prove $S_t \leq \sqrt{\frac{t}{\epsilon}} \log^c n$ for some constant $c > 0$. This can also be proven by the concentration inequality. To prove item 4a, we only need to utilize the fact that for any $t < 1/(\mu^2 t)$, $\Pr[S_t = x] = O(\frac{1}{\min\{\sqrt{t}, \sqrt{n-t}\}})$. Our

communication complexity at time $t = \tilde{O}(1/(\mu^2\epsilon))$ is

$$\Pr[S_t \leq \sqrt{k}/\epsilon] + k \Pr[S_t \geq \sqrt{k}/\epsilon] = O\left(\frac{\sqrt{k}}{\epsilon} \frac{1}{\min\{\sqrt{t}, \sqrt{n-t}\}}\right)$$

which implies that the expected communication cost in the first phase is $\tilde{O}(\sqrt{\frac{k}{t\epsilon}})$.

The communication in the second phase can be directly implied from Huang et al. [61].

□

3.3.4 Fractional Brownian Motion

In this section we consider the counting process S_t evolving as a fractional Brownian motion with parameters $\sigma > 0$ and $0 < H < 1$ where we extend the counting process to continuous time in a natural manner. We briefly discuss some of the basic properties of fractional Brownian motion (more details can be found, e.g. in [95]). Fractional Brownian motion is a process with stationary increments whose finite dimensional distributions are Gaussian. Specifically, for a fractional Brownian motion S_t , we have $E[S_t] = 0$, for every $t \geq 0$ and the covariance of the process is defined as

$$E[S_t S_u] = \frac{\sigma^2}{2}(|t|^{2H} + |u|^{2H} - |u - t|^{2H}).$$

Thus, the variance of the process is $E[S_t^2] = \sigma^2|t|^{2H}$, for every $t \geq 0$. The parameter H is known as the Hurst parameter. For $H = 1/2$, the process corresponds to a Brownian motion whose increments are independent. For $0 < H < 1/2$, the variance of S_t grows sublinearly with t and the process has a negative autocorrelation while for $1/2 < H < 1$, the variance of S_t grows superlinearly with t . The process is self-similar, meaning that random variables S_{at} and $a^H S_t$ have the same distribution. To simplify notation, in the remainder, we will assume $\sigma^2 = 1$. Notice that this is without loss of

generality as it amounts only to rescaling of the time units. It is noteworthy that the fractional Brownian motion is one of standard statistical models that captures some of the salient properties of temporal statistical dependencies that were observed in many natural phenomena, including self-similarity and long-range dependency (see, e.g. [95]).

We present an algorithm that requires only an upper bound on the Hurst parameter H and guarantees continual tracking within prescribed relative accuracy with high probability for the range $H \in [1/2, 1)$. Note that this is the range of particular interest in practice since typical values of the Hurst parameter observed in nature fall precisely in this region. For the purpose of deriving an upper bound on the communication complexity, we will write the sampling probability in the following form, for $1 < \delta \leq 2$,

$$\text{Sample-Prob}(S_t, t) = \min \left\{ \frac{\alpha_\delta \log^{1+\delta/2} n}{(\epsilon |S_t|)^\delta}, 1 \right\} \quad (3.30)$$

where $\alpha_\delta = c(2(c+1))^{\delta/2}$, for any $c > 3/2$.

As before, we start with the single site ($k=1$) case. We have the following theorem.

Theorem 3.3.12. *For the single site ($k = 1$) case, the randomized algorithm Non-monotonic Counter with the sampling probability as in (3.30) guarantees to track the count within the relative accuracy $\epsilon > 0$ with probability $1 - 1/n$ for every $1/2 \leq H \leq 1/\delta$, where $1 < \delta \leq 2$, and uses the total expected communication of $O(n^{1-H}/\epsilon \cdot \log^{1/2+1/\delta} n)$ messages.*

We observe that for standard Brownian motion, which we may interpret as a continuous-time analog of a random walk, we have $H = 1/\delta = 1/2$, and in this

case, the sampling probability and the result of the last theorem matches that of Theorem 3.3.1. For values of the Hurst parameter H in $(1/2, 1)$, the communication complexity of the algorithm is sublinear in n , with the upper bound increasing with n as a polynomial with the exponent decreasing with H as $1 - H$ (up to a polylogarithmic factor). Note that this is inline with the intuition as a larger value of the parameter H means a larger variance and thus less of a concentration around value zero where the relative error tolerance is the most stringent.

Proof of Theorem 3.3.12. We shall first analyze the communication cost. Then we argue that the algorithm is correct with high probability.

Communication cost. We first note the following lemma.

Lemma 3.3.13. *Suppose S is a Gaussian random variable with mean zero and variance $\sigma^2 > 0$. Then, for every constants $c > 0$ and $\delta > 1$,*

$$\mathbb{E}[\min\{c|S|^{-\delta}, 1\}] \leq \sqrt{\frac{2}{\pi}} \frac{\delta}{\delta - 1} \frac{c^{1/\delta}}{\sigma}.$$

Proof. We need to show that for a Gaussian random variable with mean zero and variance $\sigma^2 > 0$, the following holds for every $c > 0$ and $\delta > 1$,

$$\mathbb{E}[\min\{c|S|^{-\delta}, 1\}] \leq \sqrt{\frac{2}{\pi}} \frac{\delta}{\delta - 1} \frac{c^{1/\delta}}{\sigma}.$$

Note that

$$\begin{aligned} \mathbb{E}[\min\{c|S|^{-\delta}, 1\}] &= \mathbb{E}[c|S|^{-\delta} I(|S|^\delta > c)] + \Pr[|S|^\delta \leq c] \\ &= 2c\mathbb{E}[S^{-\delta} I(S > c^{1/\delta})] + 2\Pr[S \in [0, c^{1/\delta}]] \\ &= \frac{2c}{\sigma^\delta} \mathbb{E}\left[\frac{1}{N} I(N > c^{1/\delta}/\sigma)\right] + 2\Pr[N \in [0, c^{1/\delta}/\sigma]] \end{aligned}$$

where N is a standard normal random variable (with mean 0 and variance 1). Now, note

$$\begin{aligned}
\frac{2c}{\sigma^\delta} \mathbb{E}\left[\frac{1}{N} I(N > c^{1/\delta}/\sigma)\right] &\leq \frac{2c}{\sigma^\delta} \frac{1}{\sqrt{2\pi}} \int_{c^{1/\delta}/\sigma}^{\infty} \frac{1}{x^\delta} e^{-\frac{x^2}{2}} dx \\
&\leq \frac{2c}{\sigma^\delta} \frac{1}{\sqrt{2\pi}} \int_{c^{1/\delta}/\sigma}^{\infty} \frac{1}{x^\delta} dx \\
&= \frac{2c}{\sigma^\delta} \frac{1}{\sqrt{2\pi}} \frac{1}{\delta-1} \left(\frac{\sigma}{c^{1/\delta}}\right)^{\delta-1} \\
&= \sqrt{\frac{2}{\pi}} \frac{1}{\delta-1} \frac{c^{1/\delta}}{\sigma}.
\end{aligned}$$

It also holds

$$2 \Pr[N \in [0, c^{1/\delta}/\sigma]] \leq 2 \cdot \frac{1}{\sqrt{2\pi}} \frac{c^{1/\delta}}{\sigma} = \sqrt{\frac{2}{\pi}} \frac{c^{1/\delta}}{\sigma}.$$

Summing the two together we complete the proof. \square

Applying the lemma to the sampling probability given in (3.30), we have for every $1 \leq t \leq n$,

$$\mathbb{E}[R_t] \leq \sqrt{\frac{2}{\pi}} \frac{\delta}{\delta-1} \frac{[\alpha(2(\alpha+1))^{\delta/2} \log^{1+\delta/2} n]^{1/\delta}}{\epsilon} \frac{1}{t^H}.$$

Therefore, the expected communication cost is

$$\sum_{t \leq n} \mathbb{E}[R_t] = O(n^{1-H}/\epsilon \cdot \log^{1/2+1/\delta} n).$$

Correctness. We next bound the probability of error. Note

$$\begin{aligned}
\Pr[V_t \leq r \mid S_t = s] &= \Pr\left[\max_{i=t+1, \dots, t+r} S_i - S_t \geq \left\lceil \frac{s}{1-\epsilon} \right\rceil - s \mid S_t = s\right] \\
&\leq n \Pr\left[S_{t+r} - S_t \geq \frac{\epsilon}{1-\epsilon} s \mid S_t = s\right]
\end{aligned} \tag{3.31}$$

We shall make use of the following lemma.

Lemma 3.3.14. *Let (Z, Y) be a bivariate Gaussian random variable with mean $(0, 0)$ and covariance matrix $\Sigma = \begin{pmatrix} \sigma_Z^2 & c_{Z,Y} \\ c_{Z,Y} & \sigma_Y^2 \end{pmatrix}$. Then, conditional on $Z \in [z, z + dz)$, Y is a Gaussian random variable with mean $\mu_{Y|z} = c_{Z,Y}/\sigma_Z^2 \cdot z$ and variance $\sigma_{Y|z}^2 = \sigma_Y^2 - c_{Z,Y}^2/\sigma_Z^2$.*

Proof. Let us denote with $\phi_{Z,Y}(z, y)$ the density of a bivariate normal random variable with mean $(0, 0)$ and covariance matrix Σ and let $\phi_Z(z)$ be the density of a normal random variable with mean zero and variance σ_Z^2 .

We have

$$\begin{aligned} \Pr[Y > y \mid Z \in [z, z + dz)] &= \frac{\Pr[Y > y, Z \in [z, z + dz)]}{\Pr[Z \in [z, z + dz)]} \\ &= \frac{1}{\phi_Z(z)} \int_y^\infty \phi_{Z,Y}(z, u) du. \end{aligned}$$

Therefore,

$$\Pr[Y > y \mid Z \in [z, z + dz)] = \frac{1}{\frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{z^2}{2\sigma_Z^2}}} \cdot \frac{1}{2\pi\sqrt{|\Sigma|}} \int_y^\infty \exp\left(-\frac{1}{2} \begin{pmatrix} z & u \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z \\ u \end{pmatrix}\right) du$$

Now, note

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} \sigma_Y^2 & -c_{Z,Y} \\ -c_{Z,Y} & \sigma_Z^2 \end{pmatrix}$$

where $|\Sigma| = \sigma_Y^2 \sigma_Z^2 - c_{Z,Y}^2$. It is readily obtained that

$$\begin{pmatrix} z & y \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z \\ y \end{pmatrix} = \frac{z^2}{\sigma_Z^2} + \frac{\sigma_Z^2}{|\Sigma|} \left(y - \frac{c_{Z,Y}}{\sigma_Z^2} z\right)^2$$

and, thus,

$$\begin{aligned}
& \frac{1}{2\pi\sqrt{|\Sigma|}} \int_y^\infty \exp\left(-\frac{1}{2} \begin{pmatrix} z & u \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z \\ u \end{pmatrix}\right) du \\
&= e^{-\frac{z^2}{2\sigma_Z^2}} \cdot \int_y^\infty \exp\left(-\frac{1}{2} \frac{\sigma_Z^2}{|\Sigma|} \left(u - \frac{c_{Z,Y}}{\sigma_Z^2} z\right)^2\right) du \\
&= e^{-\frac{z^2}{2\sigma_Z^2}} \cdot \frac{1}{2\pi\sigma_Z} \int_{\frac{\sigma_Z}{\sqrt{|\Sigma|}}(y - \frac{c_{Z,Y}}{\sigma_Z^2} z)}^\infty e^{-\frac{1}{2} w^2} dw \\
&= e^{-\frac{z^2}{2\sigma_Z^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_Z} \bar{\Phi}\left(\frac{\sigma_Z}{\sqrt{|\Sigma|}} \left(y - \frac{c_{Z,Y}}{\sigma_Z^2} z\right)\right)
\end{aligned}$$

where in the second equation we use the change of variable $w = \frac{\sigma_Z}{\sqrt{|\Sigma|}} \left(u - \frac{c_{Z,Y}}{\sigma_Z^2} z\right)$.

Therefore, we obtain

$$\Pr[Y > y \mid Z \in [z, z + dz]] = \bar{\Phi}\left(\frac{\sigma_Z}{\sqrt{|\Sigma|}} \left(y - \frac{c_{Z,Y}}{\sigma_Z^2} z\right)\right).$$

□

Let G_r be a Gaussian random variable with mean zero and variance

$$\sigma_G^2\left(\frac{s}{b}\right) = \frac{\sigma_{S_{t+r}|s}^2}{\left(1 - \frac{\mu_{S_{t+r}|s}}{b}\right)^2}$$

then, note $\Pr[S_{t+r} > b \mid S_t \in [s, s + ds]] = \Pr[G_r > b]$.

Defining $\kappa_G = \mathbb{E}[S_t(S_t - S_{t+r})]$ and $\delta_G^2 = \mathbb{E}[(S_{t+r} - S_t)^2]$, it can be readily showed that

$$\sigma_G^2\left(\frac{s}{b}\right) = \frac{\delta_G^2 - \kappa_G^2 / \sigma_{S_t}^2}{\left(1 - \frac{s}{b} + \frac{s}{b} \kappa_G / \sigma_{S_t}\right)^2}.$$

Now, since $(S_u, u \geq 0)$ is a fractional Brownian motion with parameter $0 < H < 1$, we have

$$\begin{aligned}
\delta_G^2 &= r^{2H} \\
\kappa_G &= \frac{1}{2}[(t+r)^{2H} - t^{2H} - r^{2H}].
\end{aligned}$$

In particular, for standard Brownian motion, $\delta_G^2 = r$ and $\kappa_G = 0$.

For given $\eta_n > 0$, $\Pr[G_r > b] \leq \eta_n$ is equivalent to

$$\sigma_G^2\left(\frac{s}{b}\right) \leq \left(\frac{b}{\bar{\Phi}^{-1}(\eta_n)}\right)^2 \quad (3.32)$$

where $\bar{\Phi}(\cdot)$ is the complementary distribution of a normal random variable.

The following is a key lemma that enables us to use a stochastic comparison to derive an algorithm that does not require exact knowledge of the Hurst parameter H , for the case $1/2 \leq H < 1$, but only an upper bound on it.

Lemma 3.3.15. *It holds $\sigma_G^2(\frac{s}{b}) \leq \frac{\delta_G^2}{(1-\frac{s}{b})^2}$, for $1/2 \leq H < 1$.*

Proof. This is evident from the fact $\kappa_G \geq 0$ iff $1/2 \leq H < 1$ where equality holds for $H = 1/2$. \square

Furthermore, we may use the fact $\bar{\Phi}(x) \leq e^{-\frac{x^2}{2}}$, for large enough x , which yields $\bar{\Phi}^{-1}(\eta_n) \leq (2 \log(1/\eta_n))^{1/2}$, for large enough n . Hence, we obtain that for (3.32) to hold for $\eta_n = 1/n^{\alpha+1}$, it suffices that $\frac{\sigma_G^2}{(1-\frac{s}{b})^2} \leq (\frac{b}{\alpha \log n})^2$, i.e.

$$r \leq \left(\frac{|b-s|}{((\alpha+1) \log n)^{1/2}}\right)^{1/H}. \quad (3.33)$$

If this condition holds for $b = s/(1-\epsilon)$, then this implies $\Pr[S_{t+r} - S_t \geq \frac{\epsilon}{1-\epsilon}s \mid S_t = s] \leq 1/n^{\alpha+1}$, and thus in view of (3.31), it follows $\Pr[V_t \leq r \mid S_t \in [s, s+ds)] \leq 1/n^\alpha$.

The rest of the proof follows by the same arguments as in the proof of Theorem 3.3.3, using $p_t = \min\{\frac{\alpha \log n}{(1-\epsilon)^2 r}, 1\}$ and r defined by taking equality in (3.33) with $b - s = \epsilon s/(1-\epsilon)$, which yields the sampling probability

$$p_t \geq \min \left\{ \frac{(2(\alpha+1) \log n)^{1/(2H)}}{(\epsilon |S_t|)^{1/H}} \cdot \alpha \log n, 1 \right\}.$$

The right-hand side is increasing with H , hence it suffices to set the sampling probability such that equality holds in the last relation with H replaced by $1/\delta$. This is our sampling probability defined in (3.30).

Finally, we bound the error probability. We showed that $\Pr[V_t < U_t \mid S_t \in [s, s + ds)] \leq 3/n^\alpha$, for every real value s and $1 \leq t \leq n$. Hence,

$$\Pr[E_n > 0] \leq \sum_{t \leq n} \mathbb{E}[R_t] \cdot O(1/n^\alpha) = O(n^{1-H-\alpha}/\epsilon \cdot \log^{1/2+1/\delta}).$$

Hence, no error occurs with probability at least $1 - 1/n$, for any $\alpha > 2 - H$. \square

Multiple sites. Finally, we look at the case with multiple sites. Let the sampling probability be as in (3.30) but with constant $\gamma_{\alpha,\delta}$ redefined as follows $\alpha_\delta = 9 \cdot 2^{\delta/2}(c + 1)^{1+\delta/2}$, for any $c > 3/2$. Using the same coupling argument as in Theorem 3.3.5, we have the following corollary.

Corollary 3.3.16. *The randomized algorithm Non-monotonic Counter, with the sampling probability given in (3.1), guarantees to track the count across k sites within the relative accuracy $\epsilon > 0$ with probability $1 - O(1/n)$ for every $1/2 \leq H \leq 1/\delta$, where $1 < \delta \leq 2$, and uses the total expected communication of $\tilde{O}(n^{1-H} k^{\frac{3-\delta}{2}}/\epsilon)$ messages.*

Proof of Corollary 3.3.16. We can use exactly the same coupling argument as in Theorem 3.3.3. The only difference is in calculating the communication costs at different stages, expressed in (3.11). The communication in the straightforward stage is $\tilde{O}(n^{1-H} \sqrt{k} \epsilon)$. To analyze the broadcast stage, let us write the sampling probability in (3.30) as $p_t = \max\{\nu/|S_t|^\delta, 1\}$ with $\nu = (9 \cdot 2^{\delta/2}(c + 1)^{1+\delta/2})^{1/\delta}/(\epsilon/\sqrt{k})$ (we use the same re-scaling of the parameter ϵ to ϵ/\sqrt{k} as in the proof of Theorem 3.3.5). Notice that $\nu^{1/\delta} < \sqrt{k}/(\epsilon\sqrt{1+\epsilon})$, and hence the probability of sending a message at time

t when in broadcast stage is $k\nu\mathbb{E}[|S_t|^{-\delta}I(|S_t| > \frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}})]$. From here, using a similar derivation as in Lemma 3.3.13, we get that the overall expected cost in the broadcast stage is $\tilde{O}(n^{1-H}k^{\frac{3-\delta}{2}}/\epsilon)$, which is the dominant cost for our choice of $1 < \delta \leq 2$. In the remainder of this proof we calculate upper bounds on the expected communication cost.

We start by considering Eq. (3.11). Let us first calculate the expected number of messages in the straightforward stage. We have

$$\Pr[S_t \leq \frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}}] = \Pr[N \leq \frac{\sqrt{k}}{t^H\epsilon\sqrt{1-\epsilon}}] \leq \sqrt{\frac{2}{\pi}} \frac{\sqrt{k}}{t^H\epsilon\sqrt{1-\epsilon}}$$

where N is a standard normal random variable.

Next, we calculate the expected number of messages in the broadcast case. Let $\nu = \frac{9(\alpha+1)(2(\alpha+1))^{\delta/2} \log^{1+\delta/2} n}{\epsilon^\delta}$, so the sampling probability is $p_t = \max\{\nu/|S_t|^\delta, 1\}$. Notice that $\nu^{1/\delta} \leq \sqrt{k}/(\epsilon\sqrt{1+\epsilon})$. We then have

$$\begin{aligned} k\mathbb{E}[R_{m,t}I(S_t \geq \frac{\sqrt{k}}{\epsilon\sqrt{1+\epsilon}})] &= k\mathbb{E}[\max\{\nu/|S_t|^\delta, 1\}I(S_t \geq \frac{\sqrt{k}}{\epsilon\sqrt{1+\epsilon}})] \\ &= k\nu\mathbb{E}[|S_t|^{-\delta}I(|S_t| > \frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}})] \\ &= k\frac{\nu}{t^{\delta H}}\mathbb{E}[\frac{1}{N}I(|N| \geq \frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}}\frac{1}{t^H})] \\ &\leq k\frac{\nu}{t^{\delta H}}\sqrt{\frac{2}{\pi}}\frac{1}{\delta-1}\left(\frac{\sqrt{k}}{\epsilon\sqrt{1-\epsilon}}\frac{1}{t^H}\right)^{-\delta+1} \\ &= \tilde{O}\left(\frac{1}{t^H}\frac{1}{\epsilon}k^{\frac{3-\delta}{2}}\right). \end{aligned}$$

We see that the broadcast stage is dominant for $1 < \delta \leq 2$, and hence the expected communication cost is $\tilde{O}(n^{1-H}k^{\frac{3-\delta}{2}}/\epsilon)$. \square

3.4 Lower Bounds

In this section, we establish matching lower bounds for the two cases of inputs: i.i.d. Bernoulli and random permutation. Recall that we denote with M_n the number of messages exchanged over an input of size n . We are interested in the lower bounds on the expected number of messages $E[M_n]$ that is necessary to track the value over an interval of n updates within ϵ relative accuracy with high probability. We use sample-path arguments to prove the results.

We start by presenting lower bounds for the single site case, first without and then with a drift. We then provide our main results that provides a lower bound parameterized with the number of sites k for the case without drift. We conclude by giving a lower bound for the case with random permutation input stream.

Theorem 3.4.1. *Consider the single site ($k = 1$) continual count-tracking problem for an input of n random i.i.d. updates without a drift ($\Pr[X_i = 1] = \Pr[X_i = -1] = 1/2$) within relative accuracy $\epsilon > 0$ with probability at least $1 - O(1/n)$. Then the expected number of messages exchanged is $\Omega(\min\{\sqrt{n}/\epsilon, n\})$.*

The key idea of the proof is the observation that whenever the value of the counter is in $\mathcal{E} = \{s \in \mathbb{Z} : |s| \leq 1/\epsilon\}$, the site must report the value to the coordinator as otherwise an error would occur with a constant probability. The proof then follows by noting that $\sum_{t \leq n} \Pr[S_t \in \mathcal{E}] = \Omega(\sqrt{n}/\epsilon)$.

Proof of Theorem 3.4.1. Let $\mathcal{E} = \{s \in \mathbb{Z} : |s| \leq 1/\epsilon\}$. Our crucial observation here is that whenever S_t walks inside the region \mathcal{E} we have $\epsilon|S_t| < 1$ and no errors are allowed. Specifically, let I_t be the indicator random variable that sets to 1 if and

only if $S_t \in \mathcal{E}$. Notice that $\mathbb{E}[I_t] = \Pr[S_t \in \mathcal{E}] = \Omega(|\mathcal{E}|/\sqrt{t}) = \Omega(1/(\sqrt{t}\epsilon))$ and $\mathbb{E}[\sum_{t \leq n} I_t] = \Theta(\min\{\sqrt{n}/\epsilon, n\})$. On the other hand, our error requirement gives us $\Pr[M_n \geq \sum_{t \leq n} I_t] \geq 1 - 1/n$. We can then derive $\mathbb{E}[M_n]$ from $\mathbb{E}[\sum_{t \leq n} I_t]$ using the following argument. Let \mathcal{A} be the subset of the probability space where $M_n \geq \sum_{t \leq n} I_t$ and let $\neg\mathcal{A}$ be the subset where this does not hold. We have

$$\begin{aligned} \mathbb{E}[M_n] &\geq \int_{\mathcal{A}} M_n dF \geq \int_{\mathcal{A}} \sum_{t \leq n} I_t dF \\ &= \mathbb{E}[\sum_{t \leq n} I_t] - \int_{\neg\mathcal{A}} \sum_{t \leq n} I_t dF \geq \mathbb{E}[\sum_{t \leq n} I_t] - 1 \end{aligned}$$

where the last equality follows from the facts that $\sum_{t \leq n} I_t \leq n$ by construction, and that $\int_{\neg\mathcal{A}} dF \leq 1/n$. \square

The lower bound in Theorem 3.4.1 is established by counting the average number of visits to the set \mathcal{E} , and we can use the same argument to establish a lower bound for the general case of Bernoulli updates with an arbitrary drift $-1 < \mu < 1$ (that is $0 < p < 1$). Intuitively, the no drift case should be the worst case with respect to communication complexity as observed in Chapter 3.3. Also, for any constant $\mu > 0$ we expect to have the lower bound similar to the bound from [61] for a monotonic counter $\mathbb{E}[M_n] = \Omega(1/\epsilon)$. It is thus of interest to ask what the lower bound would be for small but non-zero drift $\mu = o(1)$. We have the following result.

Theorem 3.4.2. *Consider the single site ($k = 1$) continual count-tracking with Bernoulli random walk updates with drift $\mu = o(1)$ and relative accuracy parameter $\epsilon > 0$. Suppose $\epsilon = \omega(1/\sqrt{n})$ and $|\mu| = O(\epsilon)$. Then, for the tracking to succeed with probability at least $1 - O(1/n)$, the expected number of messages is $\Omega\left(\min\left\{\sqrt{n}, \frac{1}{|\mu|}\right\} \cdot \frac{1}{\epsilon}\right)$.*

The result is in line with the intuition that any non-zero drift may only reduce the communication complexity, and it matches the bound in [61] for large enough μ . Our lower bound matches the corresponding upper bound result (presented in Theorem 3.3.4) up to poly-logarithmic factors.

Proof of Theorem 3.4.2. The proof is by direct analysis of the probability of event $S_t \in \mathcal{E} = \{s \in Z : |s| \leq 1/\epsilon\}$, where the distribution of S_t is given by

$$\Pr[S_t = s] = \binom{t}{\frac{t+s}{2}} p^{\frac{t+s}{2}} (1-p)^{\frac{t-s}{2}}.$$

We remark that in the proof it is implicitly assumed that p , μ and ϵ are sequences indexed with n , but we omit to make this explicit in the notation for simplicity of presentation.

For convenience, we introduce the notation $\sigma^2 = \text{Var}[X_1] = 4p(1-p)$ and let $\rho = \sqrt{\frac{p}{1-p}}$. We then have

$$\Pr[S_t = s] = \sigma^t \frac{1}{2^t} \binom{t}{\frac{t+s}{2}} \rho^s.$$

Since $\frac{1}{2^t} \binom{t}{\frac{t+s}{2}} = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{t}}$ for $s = o(\sqrt{t})$, we have

$$\Pr[S_t = s] = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{t}} \sigma^t \rho^s \cdot [1 + o(1)], \text{ for } s = o(\sqrt{t}).$$

In order to simplify the notation and with a slight abuse in the remainder of the proof we omit to write the factor $[1 + o(1)]$.

Let $\theta_0 \geq 0$ and $\theta_1 \geq 0$ be such that $|\theta_0| = o(\sqrt{t})$ and $\theta_1 = o(\sqrt{t})$ and consider $\Pr[S_t \in [-\theta_0, \theta_1]]$, for $t = 1, 2, \dots, n$. For $1/2 < p < 1$ and $s = o(\sqrt{t})$, we have

$$\begin{aligned} \Pr[S_t \in [-\theta_0, \theta_1]] &= \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{t}} \sigma^t \sum_{s=-\theta_0}^{\theta_1} \rho^s \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{t}} \sigma^t \left(\frac{\rho^{\theta_1+1} - 1 + \rho^{\theta_0+1} - 1}{\rho - 1} - 1 \right). \end{aligned}$$

Let $E_n[-\theta_0, \theta_1]$ denote the number of visits of the set $[-\theta_0, \theta_1]$ by the counter S_t and let $\tau_n = \omega(\max\{\theta_0, \theta_1\})$. Then, note

$$\mathbb{E}[E_n[-\theta_0, \theta_1]] \geq \sum_{t=\tau_n}^n \Pr[S_t \in [-\theta_0, \theta_1]] = \left(\frac{\rho^{\theta_1+1} - 1 + \rho^{\theta_0+1} - 1}{\rho - 1} - 1 \right) \cdot \sqrt{\frac{2}{\pi}} \sum_{t=\tau_n}^n \frac{1}{\sqrt{t}} \sigma^t.$$

Notice that for every $c > 0$,

$$\sum_{t=\tau_n}^n \frac{1}{\sqrt{t}} e^{-ct} = \int_{\tau_n}^n \frac{1}{\sqrt{t}} e^{-ct} dt \geq \int_{2\sqrt{\tau_n}}^{2\sqrt{n}} e^{-\frac{c}{4}u^2} du = 2\sqrt{\frac{\pi}{c}} [\Phi(\sqrt{2cn}) - \Phi(\sqrt{2c\tau_n})]$$

where Φ is the distribution of a standard normal random variable.

Therefore,

$$\mathbb{E}[E_n[-\theta_0, \theta_1]] \geq \frac{4a_n b_n}{\log^{1/2}(\frac{1}{\sigma^2})} \quad (3.34)$$

where

$$\begin{aligned} a_n &= \frac{\rho^{\theta_1+1} - 1 + \rho^{\theta_0+1} - 1}{\rho - 1} - 1 \\ b_n &= \Phi(\log^{1/2}(\frac{1}{\sigma^2})\sqrt{n}) - \Phi(\log^{1/2}(\frac{1}{\sigma^2})\sqrt{\tau_n}). \end{aligned}$$

Now, we consider the case of a small but non-zero drift $\mu = p - (1 - p) = o(1)$ and $\theta_0 = \theta_1 = 1/\epsilon$ where $1/\epsilon$ is a positive integer. We will assume that $\tau_n = o(n)$ and $\tau_n = \omega(1/\epsilon^2)$, thus $\epsilon = \omega(1/\sqrt{n})$.

It is straightforward to show that the following asymptotes hold:

$$\begin{aligned} \rho &= 1 + \mu + O(\mu^2) \\ \rho - 1 &= \mu + O(\mu^2) \\ \sigma^2 &= 1 - \mu^2 \\ \log(\frac{1}{\sigma^2}) &= \mu^2 + O(\mu^3) \end{aligned}$$

For the term a_n , it holds

$$a_n = 2 \frac{\rho^{\frac{1}{\epsilon}+1} - 1}{\rho - 1} - 1 = \frac{2}{\mu} (e^{\frac{\mu}{\epsilon}} - 1) \cdot [1 + o(1)].$$

Hence, $a_n = \Theta(1/\mu)$, for $\mu = O(\epsilon)$. Notice that for the case $\epsilon = o(\mu)$, a_n grows as $\Omega(e^{\mu/\epsilon})$.

For the term b_n , we observe

$$\begin{aligned} b_n &= \Phi(\log^{1/2}(\frac{1}{\sigma^2})\sqrt{n}) - \Phi(\log^{1/2}(\frac{1}{\sigma^2})\sqrt{\tau_n}) \\ &= [\Phi(\mu\sqrt{n}) - \Phi(\mu\sqrt{\tau_n})] \cdot [1 + o(1)] \end{aligned}$$

and is easy to derive that $b_n = \Theta(1)$, for $\mu = O(1/\sqrt{n})$ and $b_n = \Theta(\mu\sqrt{n})$ for $\mu = o(1/\sqrt{n})$. Indeed, these are easy to derive from the above asymptotes and the facts $\Phi(\mu\sqrt{n}) - \Phi(\mu\sqrt{\tau_n}) = 1 - 1/2 = 1/2$ for $\mu = \omega(1/\sqrt{n})$ and $\Phi(\mu\sqrt{n}) - \Phi(\mu\sqrt{\tau_n}) \geq \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2 n}{2}} \mu(\sqrt{n} - \sqrt{\tau_n})$.

The assertion of the theorem follows by plugging the derived asymptotes for a_n , b_n and $\log^{1/2}(1/\sigma^2) = \mu[1 + o(1)]$ into (3.34).

□

We now move to the main result of this section which provides a lower bound that is parameterized with the number of sites k . We consider only the non-drift case, as this is used to establish the lower bound for the permutation model. While the proof for $k = 1$ case essentially only needs to exploit the structure of a simple random walk, here we need to carefully integrate techniques from communication complexity theory with the structure of random walks. The main step is a reduction to a query problem (Lemma 3.4.4) that at a time instance asks whether the sum of updates over all k sites is larger than or equal to a $\Theta(\sqrt{k})$ threshold, which requires

$\Omega(k)$ communication to guarantee a sufficiently low probability of error; otherwise, the overall error probability does not satisfy the requirements.

We start our analysis by introducing a building block for communication complexity.

Definition 3.4.3 (Tracking k inputs). *Let c be a constant. Consider the following functional monitoring problem: let X_1, X_2, \dots, X_k be i.i.d. variables from $\{-1, 1\}$ such that $\Pr[X_i = -1] = \Pr[X_i = 1] = 1/2$ that arrive uniformly to each of the sites (i.e. each site receives exactly one update). Upon the arrival of the last update, we require the coordinator to*

- *be able to tell whether the sum is positive or negative if $|\sum_{i \leq k} X_i| \geq c\sqrt{k}$.*
- *do anything (i.e. no requirement) when $|\sum_{i \leq k} X_i| < c\sqrt{k}$.*

We have the following lemma.

Lemma 3.4.4. *Solving the tracking k inputs problem with probability $1 - c_0$ (w.r.t. both the protocol and the input) for some constant c_0 requires $\Theta(k)$ communication.*

Proof of Lemma 3.4.4. Our proof follows that of Lemma 2.2 in [61]. Here, we only need to argue that the communication lower bound still holds for a two round *deterministic* protocol such that

- in the first round, a subset of sites report their individual values to the coordinator;
- in the second round, the coordinator probes a subset of sites to make the decision.

The lemma follows in view of the fact that a randomized protocol can be seen as a distribution over a set of deterministic algorithms. It suffices to consider the case where $o(k)$ messages are sent in the first round, as otherwise we are done with the proof. This essentially reduces the communication complexity problem to a known sampling complexity problem [61]. The only remaining obstacle here is that the input distribution under our consideration is not exactly the same as the one studied in [61]. Therefore, we need to re-establish the sampling lower bound in our setting, which is provided in the remainder of this section.

Let $k' = \Theta(k)$ be the number of sites that have not sent any messages in the first round, and without loss of generality, assume that these sites are $1, 2, \dots, k'$. Since the number of messages sent in the first round is $o(k)$, in the second round, we need to solve a problem that is at least as hard as the following one:

- answer whether the sum $\sum_{i \leq k} X_i$ is positive or negative, if $|\sum_{i \leq k'} X_i| \geq c\sqrt{k'}$;
- do anything (i.e. no requirement) when $|\sum_{i \leq k'} X_i| < c\sqrt{k'}$

where c is a positive constant.

Let us denote with z the number of sites that are sampled by the coordinator in the second round, and without loss of generality, let us assume that these sites are $1, 2, \dots, z$. To contradict, let us suppose $z = o(k')$. Let $N = \sum_{i \leq z} X_i$ be the cumulative update value of the sampled sites and $U = \sum_{z < i \leq k'} X_i$ be the cumulative update value of the unsampled sites. Clearly, the optimal detection algorithm for the sampling problem is to declare $\sum_{i \leq k'} X_i > c\sqrt{k'}$ if $N > 0$, to declare $\sum_{i \leq k'} X_i < -c\sqrt{k'}$ if $N < 0$ and to declare either (with probability $1/2$) if $N = 0$. The probability

of error is then

$$\begin{aligned} \Pr[\text{error}] &\geq \Pr[N < 0, N + U \geq c\sqrt{k'}] \\ &\geq \Pr[-c\sqrt{z} \leq N < 0] \Pr[U \geq c(\sqrt{k'} + \sqrt{z})]. \end{aligned}$$

Since N is a sum of independent and identically distributed random variables of mean zero and variance 1, we have $E[N] = 0$ and $\text{Var}[N] = z$, and thus $\Pr[-c\sqrt{z} \leq N < 0] = \Theta(1)$. Similarly, since U is a sum of independent and identically random variables of mean zero and variance 1, we have $E[U] = 0$ and $\text{Var}[U] = k' - z$, and under our assumption $z = o(k')$, it holds $c(\sqrt{k'} + \sqrt{z}) = c\sqrt{k' - z} \cdot [1 + o(1)] = c\text{Var}[U] \cdot [1 + o(1)]$, and thus $\Pr[U \geq c(\sqrt{k'} + \sqrt{z})] = \Theta(1)$. Therefore, the probability of error is $\Omega(1)$ which contradicts the error requirement, for sufficiently small constant c_0 in the statement of the lemma.

□

We are now ready to present our main lower bound theorem for the k -site case.

Theorem 3.4.5. *For the case of $k < n$ sites, the expected amount of communicated messages to guarantee relative accuracy $\epsilon > 0$ with probability at least $1 - O(1/n)$ is $\Omega(\min\{\sqrt{kn}/\epsilon, n\})$.*

Here, again our lower bound matches the corresponding upper bound presented in Theorem 3.3.5. The intuition behind the result is as follows. We chop the stream into phases of k updates each, where each site gets exactly one update per phase. If the counter value S_t is in between $-\sqrt{k}/\epsilon$ and \sqrt{k}/ϵ , we show that our problem is equivalent to the tracking of k input problems, and $\Theta(k)$ messages need to be sent

to guarantee correctness. Summing the expected number of visits of the counter to these states, we obtain the lower bound.

Proof of Theorem 3.4.5. We partition the updates into n/k phases, each of which consists of k updates. In each phase, the k updates are randomly matched to k sites (so that each site receives exactly one update). Let I_j be an indicator random variable that sets to 1 when, at the beginning of the j th phase, the sum is in the interval $[-a_{j,k,\epsilon}, a_{j,k,\epsilon}]$ where $a_{j,k,\epsilon} \triangleq \min\{\sqrt{k}/\epsilon, \sqrt{jk}\}$. Notice that when the sum is in the interval $[-a_{j,k,\epsilon}, a_{j,k,\epsilon}]$, the additive error we can tolerate is at most $\epsilon\sqrt{k}/\epsilon = \sqrt{k}$. Therefore, at the end of the j th stage, the tracking algorithm has to be able to tell whether the absolute value of j -th phase's sum is below $-\sqrt{k}$, above \sqrt{k} , or in between the two. This is at least as difficult as the *tracking k inputs* problem we studied above, with $\Theta(k)$ communication lower bound.

Let M_n be the total number of messages exchanged between the coordinator and the sites. Our correctness requirement gives us $\Pr[M_n \geq \Omega(k \sum_{i \leq n/k} I_i)] \geq 1 - 1/n$. Using the fact that $E[\sum_i I_i] = \min\{\sqrt{n/(\epsilon k)}, n/k\}$, and following similar arguments as in the proof of Theorem 3.4.1, we get $\Omega(\min\{\sqrt{kn}/\epsilon, n\})$.

□

Random Permutation. Finally, we have the following corollary providing a lower bound on the communication complexity for randomly permuted input stream.

Corollary 3.4.6. *The expected total communication in presence of a randomly permuted adversarial input, with $-1 \leq a_t \leq 1$ for all $1 \leq t \leq n$, is at least $\Omega(\sqrt{kn}/\epsilon)$ messages.*

Proof of Corollary 3.4.6. Consider an adversary that select each input a'_t randomly such that $\Pr[a'_t = 1] = \Pr[a'_t = -1] = 1/2$. Then the process (a_t) obtained by randomly permuting (a'_t) is also a sequence of Bernoulli variables, and from Theorem 3.4.5 we know that $E[M_n] \geq \Omega(\sqrt{kn}/\epsilon)$. Clearly, using an averaging argument, there is at least one deterministic sequence a'_t that, randomly permuted, requires on average $\Omega(\sqrt{kn}/\epsilon)$ messages. This proves the claim. \square

3.5 Applications

3.5.1 Tracking the Second Frequency Moment

We now apply our distributed counter algorithms for continuously tracking second frequency moment of the randomly ordered stream. Let us recall the F_2 problem. The input stream consists of a_1, a_2, \dots, a_n , where $a_t = (\alpha_t, z_t)$, α_t are all items from the universe $[m]$, and $z_t \in \{-1, 1\}$ for all $t \leq n$. Denote with $m_i(t) = \sum_{s \leq t: \alpha_s = i} z_s$ the sum of elements of type i in the stream at time t . Here, we allow the count $m_i(t)$ to be non-monotonic in t (i.e. allow decrements of the counts). Our goal is to continuously track the second moment of the stream, i.e. $F_2(t) = \sum_{i \leq m} m_i^2(t)$, at the coordinator. We shall refer to this problem as *monitoring $F_2(t)$ with decrements*.

Next, we review the fast AMS sketching algorithm for this problem (see [32] and references therein). Consider a set of counters $(S_{i,j})_{i \leq I, j \leq J}$ whose values at the t -th update are $S_{ij}(t)$, and set $S_{ij}(0) = 0$. Let $g_j : [m] \rightarrow \{-1, 1\}$ and $h_j : [m] \rightarrow J$ (for $j \leq I$) be two sets of 4-wise independent hash functions. Upon receiving the t -th item (α_t, z_t) , we add $z_t \cdot g_j(\alpha_t)$ to all $(S_{j, h_j(\alpha_t)}(t))_{j \leq I}$. When $I = O(\log 1/\delta)$ and

$J = O(1/\epsilon^2)$, we are able to recover F_2 with ϵ -relative guarantee with prob. $1 - \delta$ for a specific t . We can then execute $\log n$ copies of the fast AMS sketches in parallel to make sure our estimator is correct for all updates.

An important property of fast AMS sketching is that it updates only $O(\log(1/\delta) + \log n) = \tilde{O}(1)$ counters after each update. Let $n_i = \sum_{s \leq n} \mathbb{I}\{\alpha_s = i\}$ be the number of occurrences of item i in the stream until time t . Clearly, $\sum_{i \in [m]} n_i = n$. Further, let $\tilde{N}_{ij} = \{k \in [m] : h_j(k) = i\}$ be the set of items that map to the counter S_{ij} . Tracking counter S_{ij} in randomly ordered stream takes $\tilde{O}(\sqrt{k \sum_{k \in \tilde{N}_{ij}} n_k / \epsilon})$ communication. Summing over all counter, and using Jensen inequality, we get that the expected total number of communicated messages is $\tilde{O}(\sqrt{kn}/\epsilon^2)$. We also remark that the lower bound $\Omega(\sqrt{kn}/\epsilon)$ for counter in randomly ordered stream is also a lower bound for $F_2(\cdot)$ for randomly ordered streams. We may summarize the upper bound and the lower bound results in the following corollary.

Corollary 3.5.1. *The communication lower bound for tracking the frequency moment $F_2(t)$ with decrements in randomly ordered stream is $\Omega(\min\{\sqrt{kn}/\epsilon, n\})$. There exists a randomized algorithm for tracking $F_2(t)$ using the total expected communication of $\tilde{O}(\sqrt{kn}/\epsilon^2)$ messages.*

3.5.2 Bayesian Linear Regression

We next describe another application of a distributed non-monotonic counter in tracking the posterior of the coefficients in a *Bayesian linear regression*. Recall the Bayesian linear regression problem (c.f. [14]): assume we are given a set of training data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$, where \mathbf{x}_i is a row-vector $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. We

are interested in carrying out a linear regression over this data, i.e. finding a $\mathbf{w} \in \mathbb{R}^d$ such that $y = \mathbf{w}^T \cdot \mathbf{x}$ best fits the training data $\{(\mathbf{x}_t, y_t)\}_{t \leq n}$. Furthermore, we impose an initial prior knowledge over the vector \mathbf{w}_0 , and in particular we assume that it follows a multivariate Gaussian distribution $\mathbf{w}_0 \sim \mathcal{N}(\mathbf{m}_0, S_0)$. Our goal is to maintain a posterior belief over \mathbf{w}_t , as the training data $\{(\mathbf{x}_t, y_t)\}_{t \leq n}$ arrives.

In the distributed functional monitoring setting, the training data $\{(\mathbf{x}_t, y_t)\}_{t \leq n}$ arrives at different sites in a streaming fashion and the coordinator has to continuously track an approximate estimate of the mean \mathbf{m}_t and the variance S_t of \mathbf{w}_t . We assume that the training data is an arbitrary bounded sequence selected by an adversary, and randomly permuted, as in the random permutation model.

We next describe how we may use $O(d^2)$ counters to track the posterior belief. Let A_t be an $t \times d$ matrix so that the i -th column of A_t is \mathbf{x}_i . Also, denote with $\mathbf{y}_t \in \mathbb{R}^t$ a vector whose i -th component is y_i . Furthermore, let β be the inverse of the variance of the noise variable in the model, i.e., $y_t = \mathbf{w}^T \cdot A_t + \mathcal{N}(0, \beta^{-1})$. The value of β is usually assumed to be a known parameter (see Bishop [14] for a detailed description of the model). It turns out that the posterior of \mathbf{w} is also a Gaussian distribution with mean \mathbf{m}_t and variance S_t , where

$$\begin{aligned} \mathbf{m}_t &= S_t(S_0^{-1}\mathbf{m}_0 + \beta A_t^T \mathbf{y}_t) \\ S_t^{-1} &= S_0^{-1} + \beta A_t^T A_t. \end{aligned} \tag{3.35}$$

The inverse of S_t^{-1} at time t is also referred as the *precision matrix*. Observe that tracking the precision matrix S_t^{-1} as well as the vector $A_t^T \mathbf{y}$ suffices to recover the posterior structure of \mathbf{w} . Our specific goal here is to continuously track S_t^{-1} and $A_t^T \mathbf{y}_t$ by using our counter algorithm.

Upon the arrival of the $t + 1$ -st update, we have

$$S_{t+1}^{-1} = S_t^{-1} + \beta \underbrace{x_{t+1}^T x_{t+1}}_{\text{outer product of } x_{t+1}}$$

and $A_{t+1}^T \mathbf{y}_{t+1} = A_t^T \mathbf{y}_t + (y_{t+1} \times (x_{t+1})_1, y_{t+1} \times (x_{t+1})_2, \dots, y_{t+1} \times (x_{t+1})_d)^T$.

Therefore, to track $S_t^{-1} \in \mathbb{R}^{d \times d}$, it suffices to keep d^2 counters $\{C_{i,j}\}_{i,j \leq d}$ such that upon the arrival of the t -th training data, $C_{i,j} \leftarrow C_{i,j} + \beta(x_t)_i(x_t)_j$. Similarly, we may keep another d copies of counters $\{D_i\}_{i \leq d}$ to track $A_t^T \mathbf{y}_t$, where $D_i \leftarrow D_i + y_t \times (x_t)_i$ at the t -th update. Notice that here our algorithm can guarantee each entry in S_t^{-1} and $A_t^T \mathbf{y}$ has at most ϵ -relative error. The actual error of our estimate for \mathbf{m}_t , however, also depends on how sensitive of the precision matrix's inverse is when it is perturbed.

The total communication complexity using this algorithm thus is $\tilde{O}(\sqrt{knd^2}/\epsilon)$, being sublinear in the size of training data for a wide range of parameters.

3.6 Geometric Progression Search for μ

Recall that in our algorithm, we need a protocol that continuously search for an estimate $\hat{\mu}$ such that $\hat{\mu} \in [(1 - \epsilon)\mu, (1 + \epsilon)\mu]$, where ϵ is given input parameter. This protocol also guarantees that $\hat{\mu}$ is found before time $\Theta(\log n/\mu^2)$.

We now describe the implementation of the protocol. In the protocol, we define a geometric progression sequence $\{\ell_i\}_{i \geq 0}$ such that $\ell_i = (1 - \epsilon)^i$, for $i \geq 0$. Our algorithm progressively tests whether $\mu \leq \ell_i$ for $i \geq 0$. When we find an ℓ_i such that $\mu \geq \ell_i$, we only need $\tilde{O}(1/\mu^2) = \tilde{O}(1/\ell_i^2)$ samples to estimate μ (here $\tilde{O}(\cdot)$ hides polylogarithmic dependencies on n).

To decide whether $\mu \leq \ell_i$, the coordinator will probe all the sites at time $t =$

$\Theta(\log n/\ell_i^2)$ (recall that we maintain a HYZ counter to track the total number of updates). When $\mu < \ell_i$, we have $\Pr[S_t \geq \ell_{i+1} \cdot t] = O(1/n^3)$ by a Chernoff bound. When $\mu \geq \ell_i$, we then can move to estimate μ , specifically, our estimator is $\hat{\mu} = S_t/t$ and we have $\Pr[\hat{\mu} \notin [(1-\epsilon)\mu, (1+\epsilon)\mu]] = O(1/n^3)$. One may see that we are able to find an estimate of μ before time $\Theta(\log n/\mu^2)$ using such a process.

Notice that the sampled data are reused in our search process. Therefore, the error events are correlated. But since the length of the data stream is n , we may use a union bound to conclude that the probability we make at least one error in the search process is $O(1/n^2)$. Finally, our search ends before we reach $\ell_i = \tilde{\Theta}(1/n)$, which means the total number of communication is $\tilde{O}(k)$.

Chapter 4

Revisiting Chernoff-Hoeffding Bounds for Markov Chains

4.1 Background

In this chapter, we establish large deviation bounds for random walks on general (irreversible) finite state Markov chains based on mixing properties of the chain in both discrete and continuous time settings. To introduce our results we focus on the discrete time setting, which we now describe.

Let M be an ergodic Markov chain with finite state space $V = [n]$ and stationary distribution π . Let (v_1, \dots, v_t) denote a t -step random walk on M starting from a distribution φ on V . For every $i \in [t]$, let $f_i : V \rightarrow [0, 1]$ be a weight function at step i so that $\mathbb{E}_{v \leftarrow \pi}[f_i(v)] = \mu > 0$ for all i . Define the total weight of the walk (v_1, \dots, v_t) by $X \triangleq \sum_{i=1}^t f_i(v_i)$. The expected total weight of the random walk (v_1, \dots, v_t) is $\mathbb{E}[\frac{1}{t}X] \approx \mu$ as $t \rightarrow \infty$.

When the v_i 's are drawn independently according to the stationary distribution π , a standard Chernoff-Hoeffding bound says that

$$\Pr [|X - \mu t| \geq \delta \mu t] \leq \begin{cases} e^{-\Omega(\delta^2 \mu t)} & \text{for } 0 \leq \delta \leq 1, \\ e^{-\Omega(\delta \mu t)} & \text{for } \delta > 1. \end{cases}$$

However, when (v_1, \dots, v_t) is a random walk on a Markov chain M , it is known that the concentration bounds depend inherently on the mixing properties of M , that is the speed at which a random walk converges toward its stationary distribution.

Variants of Chernoff-Hoeffding bounds for random walk on Markov chains have been studied in several fields with various motivations [48, 55, 67, 75, 79, 106, 108]. For instance, these bounds are linked to the performance of Markov chain Monte Carlo integration techniques [64, 75]. They have also been applied to various online learning problem [103], testing properties of a given graph [49], leader election problems [67], analyzing the structure of the social networks [9, 85], understanding the performance of data structures [45], and computational complexity [55]. Improving such bounds is therefore of general interest.

We improve on previous work in two ways. First, all the existing deviation bounds, as far as we know, are based on the *spectral expansion* $\lambda(M)$ of the chain M . This spectral expansion $\lambda(M)$ characterizes how much M can stretch vectors in \mathbf{R}^n under a normed space defined by the stationary distribution π , which coincides with the second largest absolute eigenvalue of M when M is reversible. (A formal definition is deferred to Chapter 4.2.) The most general result for Markov chains in

this form (see, e.g. [79, 106]) is

$$\Pr [|X - \mu t| \geq \delta \mu t] \leq \begin{cases} \|\varphi\|_{\pi} e^{-\Omega((1-\lambda)\delta^2 \mu t)} & \text{for } 0 \leq \delta \leq 1, \\ \|\varphi\|_{\pi} e^{-\Omega((1-\lambda)\delta \mu t)} & \text{for } \delta > 1. \end{cases} \quad (4.1)$$

where φ is an arbitrary initial distribution and $\|\cdot\|_{\pi}$ is the π -norm (which we define formally later).

However, for general irreversible Markov chains, the spectral expansion λ does not directly characterize the mixing time of a chain and thus may not be a suitable parameter for such bounds. A Markov chain M could mix rapidly, but have a spectral expansion λ close to 1, in which case Eq. (4.1) does not yield meaningful bound. In fact there is a way to modify any given Markov chain M so that the modified Markov chain M' has (asymptotically) the same mixing-time as M , but the spectral expansion of M' equals 1 (Chapter 4.4 gives a detailed construction). It is therefore natural to seek a Chernoff-type bound for Markov chains directly parameterized by the chain's mixing time T .

Second, most previous analyses for deviation bounds such as Eq. (4.1) are based on *non-elementary* methods such as perturbation theory [48, 75, 79, 108]. Kahale [67] and Healy [55] provided two elementary proofs for reversible chains, but their results yield weaker bounds than those in Eq. (4.1). Recently, Wagner [106] provided another elementary proof for reversible chains matching the form in Eq. (4.1). Together with the technique of “reversibilization” [43, 79], Wagner’s analysis can be generalized to irreversible chains. However, his use of decoupling on the linear projections outright arguably leads to a loss of insight; here we provide an approach based on directly tracing the corresponding sequence of linear projections, in the spirit of [55]. This

more elementary approach allows us to tackle both reversible and irreversible chains in a unified manner that avoids the use of “reversibilization”.

As we describe below, we prove a Chernoff-type bound for general irreversible Markov chains with general weight functions f_i based on the standard L_1 (variation distance) mixing time of the chain, using elementary techniques based on extending ideas from [55]. The exponents of our bounds are tight up to a constant factor. As far as we know, this is the first result that shows that the mixing time is sufficient to yield these types of concentration bounds for random walks on Markov chains. Along the way we provide a unified proof for (4.1) for both reversible and irreversible chains based only on elementary analysis. This proof may be of interest in its own right.

4.2 Preliminaries

Throughout this Chapter we shall refer M as the discrete time Markov chain under consideration. Depending on the context, M shall be interpreted as either the chain itself or the corresponding transition matrix (i.e. it is an n by n matrix such that $M_{i,j}$ represents the probability a walk at state i will move to state j in the next step). For the continuous time counterpart, we write Λ as the generator of the chain and let $M(t) = e^{t\Lambda}$, which represents the transition probability matrix from t_0 to $t_0 + t$ for an arbitrary t_0 .

Let u and w be two distributions over the state space \mathbf{V} . The *total variation* distance between u and w is $\|u - w\|_{TV} = \max_{A \subseteq \mathbf{V}} \left| \sum_{i \in A} u_i - \sum_{i \in A} w_i \right| = \frac{1}{2} \|u - w\|_1$.

Let $\epsilon > 0$. The mixing time of a *discrete time* Markov chain M is $T(\epsilon) = \min \{t : \max_x \|xM^t - \pi\|_{TV} \leq \epsilon\}$, where x is an arbitrary initial distribution. The

mixing time of a *continuous time* Markov chain specified by the generator Λ is $T(\epsilon) = \min \{t : \max_x \|xM(t) - \pi\|_{TV} \leq \epsilon\}$, where $M(t) = e^{\Lambda t}$.

We next define an inner product space specified by the stationary distribution π :

Definition 4.2.1 (Inner product under π -kernel). *Let M be an ergodic Markov chain with state space $[n]$ and π be its stationary distribution. Let u and v be two vectors in R^n . The inner product under the π -kernel is $\langle u, v \rangle_\pi = \sum_{x \in [n]} \frac{u_x v_x}{\pi(x)}$.*

We may verify that $\langle \cdot, \cdot \rangle_\pi$ indeed forms an inner product space by checking it is symmetric, linear in the first argument, and positive definite. The π -norm of a vector u in R^n is $\|u\|_\pi = \sqrt{\langle u, u \rangle_\pi}$. Note that $\|\pi\|_\pi = 1$. For a vector $x \in R^n$, we write $x^\parallel = \langle x, \pi \rangle_\pi \pi$ for its component along the direction of π and $x^\perp = x - x^\parallel$ for its component perpendicular to π .

We next define the *spectral norm* of a transition matrix.

Definition 4.2.2 (Spectral norm). *Let M the transition matrix of an ergodic Markov chain. Define the spectral norm of M as $\lambda(M) = \max_{\langle x, \pi \rangle_\pi = 0} \frac{\|xM\|_\pi}{\|x\|_\pi}$.*

When M is clear from the context, we shall simply write λ for $\lambda(M)$. We shall also refer $1 - \lambda(M)$ as the *spectral gap* of the chain M . In the case when M is reversible, $\lambda(M)$ coincides with the second largest eigenvalue of M (the largest eigenvalue of M is always 1). However, when M is irreversible, such relation does not hold (one hint to realize that the eigenvalues of M for an irreversible chain can be complex, and the notion of being the second largest may not even be well defined). Nevertheless, we can still connect $\lambda(M)$ with an eigenvalue of a matrix related to M . Specifically, let \tilde{M} be the time reversal of M : $\tilde{M}(x, y) = \frac{\pi(y)M(y, x)}{\pi(x)}$. The *multiplicative reversibilization*

$R(M)$ of M is $R(M) \equiv M\tilde{M}$. The value of $\lambda(M)$ then coincides with the square root of the second largest eigenvalue of $R(M)$, i.e. $\lambda(M) = \sqrt{\lambda(R(M))}$. Finally, notice that the stationary distribution of M , \tilde{M} , and R are all the same. These facts can be found in [43].

4.3 Chernoff-Hoeffding Bounds for Discrete Time Markov Chains

We now present our main result formally.

Theorem 4.3.1. *Let M be an ergodic Markov chain with state space $[n]$ and stationary distribution π . Let $T = T(\epsilon)$ be its ϵ -mixing time for $\epsilon \leq 1/8$. Let (V_1, \dots, V_t) denote a t -step random walk on M starting from an initial distribution φ on $[n]$, i.e., $V_1 \leftarrow \varphi$. For every $i \in [t]$, let $f_i : [n] \rightarrow [0, 1]$ be a weight function at step i such that the expected weight $\mathbb{E}_{v \leftarrow \pi}[f_i(v)] = \mu$ for all i . Define the total weight of the walk (V_1, \dots, V_t) by $X \triangleq \sum_{i=1}^t f_i(V_i)$. There exists some constant c (which is independent of μ , δ and ϵ) such that*

$$\begin{aligned} 1. \Pr[X \geq (1 + \delta)\mu t] &\leq \begin{cases} c\|\varphi\|_\pi \exp(-\delta^2 \mu t / (72T)) & \text{for } 0 \leq \delta \leq 1 \\ c\|\varphi\|_\pi \exp(-\delta \mu t / (72T)) & \text{for } \delta > 1 \end{cases} \\ 2. \Pr[X \leq (1 - \delta)\mu t] &\leq c\|\varphi\|_\pi \exp(-\delta^2 \mu t / (72T)) \quad \text{for } 0 \leq \delta \leq 1 \end{aligned}$$

Before we continue our analysis, we remark on some aspects of the result.

Optimality of the bound The bound given in Theorem 4.3.1 is optimal among all bounds based on the mixing time of the Markov chain, in the sense that for any

given T and constant ε , one can find a δ , a family of functions $\{f_i : V \rightarrow [0, 1]\}$, and a Markov chain with mixing time $T(\varepsilon) = T$ that has deviation probabilities matching the exponents displayed in Theorem 4.3.1, up to a constant factor. In this regard, the form of our dependency on T is tight for constant ε . For example, consider the following Markov chain:

- The chain consists of 2 states s_1 and s_2 .
- At any time step, with probability p the random walk jumps to the other state and with probability $1 - p$ it stays in its current state, where p is determined below.
- for all f_i , we have $f_i(s_1) = 1$ and $f_i(s_2) = 0$.

Notice that the stationary distribution is uniform and $T(\epsilon) = \Theta(1/p)$ when ϵ is a constant. Thus, we shall set $p = \Theta(1/T)$ so that the mixing-time $T(\varepsilon) = T$. Let us consider a walk starting from s_1 for sufficiently large length t . The probability that the walk stays entirely in s_1 up to time t is $(1 - p)^t \approx e^{-tp} = \exp(-\Theta(t/T))$. In other words, for $\delta = 1$ we have

$$\Pr[X \geq (1+\delta)\mu t] = \Pr[X \geq t] = \Pr[\text{the walk stays entirely in } s_1] = \exp(-\Theta(t/T(\epsilon))).$$

This matches the first bound in Theorem 4.3.1 asymptotically, up to a constant factor in the exponent. The second bound can be matched similarly by switching the values of $f_i(\cdot)$ on s_1 and s_2 . Finally, we remark that this example only works for $\epsilon = \Omega(1)$, which is how mixing times appear in the usual contexts. It remains open, though, whether our bounds are still optimal when $\epsilon = o(1)$.

Dependency on the threshold ϵ of the mixing time Note that the dependence of ϵ only lies on $T(\epsilon)$. Since $T(\epsilon)$ is non-decreasing in ϵ , it is obvious that $\epsilon = 1/8$ gives the best bound in the setting of Theorem 4.3.1. In fact, a more general form of our bound, as will be seen along our derivation later, replaces $1/72$ in the exponent by a factor $(1 - \sqrt{2\epsilon})/36$. Hence the optimal choice of ϵ is the maximizer of $(1 - \sqrt{2\epsilon})/T(\epsilon)$ (with $\epsilon < 1/2$), which differs for different Markov chains. Such formulation seems to offer incremental improvement and so we choose to focus on the form in Theorem 4.3.1.

Comparison with spectral expansion based Chernoff bound The bound given in Theorem 4.3.1 is *not* always stronger than spectral expansion based Chernoff bounds (4.1) that is presented in, for example, Lezaud [79] and Wagner [106]. Consider, for instance, a random constant degree regular graph G . One can see that the spectral gap of the Markov chain induced by a random walk over G is a constant with high probability. On the other hand, the mixing time of the chain is at least $\Omega(\log n)$ because the diameter of a constant degree graph is at least $\Omega(\log n)$. Lezaud [79] or Wagner [106] gives us a concentration bound $\Pr[X \geq (1 + \epsilon)\mu t] \leq c\|\varphi\|_\pi \exp(-\Theta(\delta^2 \mu t))$ when $\delta < 1$ while Theorem 4.3.1 gives us $\Pr[X \geq (1 + \epsilon)\mu t] \leq c\|\varphi\|_\pi \exp(-\Theta(\delta^2 \mu t / (\log n)))$.

Comparison with a union bound Assuming the spectral expansion based Chernoff bound in Lezaud [79] and Wagner [106], there is a simpler analysis to yield a mixing time based bound in a similar but weaker form than Theorem 4.3.1: we first divide the random walk (V_1, \dots, V_t) into $T(\epsilon)$ groups for a sufficiently small ϵ such that the i th group consists of the sub-walk $V_i, V_{i+T(\epsilon)}, V_{i+2T(\epsilon)}, \dots$. The walk in each group

is then governed by the Markov chain $M^{T(\epsilon)}$. This Markov chain has unit mixing time and as a result, its spectral expansion can be bounded by a constant (by using our Claim 4.3.1 below). Together with a union bound across different groups, we obtain

$$\begin{aligned} 1. \Pr[X \geq (1 + \delta)\mu t] &\leq \begin{cases} cT\|\varphi\|_\pi \exp(-\delta^2\mu t/(72T)) & \text{for } 0 \leq \delta \leq 1 \\ cT\|\varphi\|_\pi \exp(-\delta\mu t/(72T)) & \text{for } \delta > 1 \end{cases} \\ 2. \Pr[X \leq (1 - \delta)\mu t] &\leq cT\|\varphi\|_\pi \exp(-\delta^2\mu t/(72T)) \quad \text{for } 0 \leq \delta \leq 1 \end{aligned} \quad (4.2)$$

Theorem 4.3.1 shaves off the extra leading factors of T in these inequalities, which has significant implications. For example, Eq. (4.2) requires the walk to be at least $\Omega(T \log T)$, while our bounds address walk lengths between T and $T \log T$. Our tighter bound further can become important when we need a tighter polynomial tail bound.

As a specific example, saving the factor of T becomes significant when we generalize these bounds to continuous-time chains using the discretization strategy in Fill [43] and Lezaud [79]. The strategy is to apply known discrete time bound on the discretized continuous time chain, say in a scale of b units of time, followed by taking limit as $b \rightarrow 0$ to yield the corresponding continuous time bound. Using this to obtain a continuous analog of Eq. (4.2) does not work, since under the b -scaled discretization the mixing time becomes T/b , which implies that the leading factor in Eq. (4.2) goes to infinity in the limit as $b \rightarrow 0$.

We now proceed to prove Theorem 4.3.1.

Proof. (of Theorem 4.3.1) We partition the walk V_1, \dots, V_t into $T = T(\epsilon)$ subgroups so that the i -th sub-walk consists of the steps (V_i, V_{i+T}, \dots) . These sub-walks can be

viewed as generated from Markov chain $N \triangleq M^T$. Also, denote

$$X^{(i)} \triangleq \sum_{0 \leq j \leq t/T} f_{i+jT}(V_{i+jT})$$

as the total weight for each sub-walk and $\bar{X} = \sum_{i=1}^T X^{(i)}/T$ as the average total weight.

Next, we follow Hoeffding's approach [57] to cope with the correlation among the $X^{(i)}$. To start,

$$\Pr[X \geq (1 + \delta)\mu t] = \Pr\left[\bar{X} \geq (1 + \delta)\frac{\mu t}{T}\right] \leq \frac{\mathbb{E}[e^{r\bar{X}}]}{e^{r(1+\delta)\mu t/T}}. \quad (4.3)$$

Now noting that $\exp(\cdot)$ is a convex function, we use Jensen's inequality to obtain

$$\mathbb{E}[e^{r\bar{X}}] \leq \sum_{i \leq T} \frac{1}{T} \mathbb{E}[e^{rX^{(i)}}]. \quad (4.4)$$

We shall focus on giving an upper bound on $\mathbb{E}[e^{rX^{(i)}}]$. This requires two steps:

- First, we show the chain N has a constant spectral gap based on the fact that it takes one step to mix.
- Second, we apply a bound on the moment generating function of $X^{(k)}$ using its spectral expansion.

Specifically, we shall prove the following claims, whose proofs will be deferred to later part of this Chapter.

Claim 4.3.1. *Let M be a general ergodic Markov chain with ϵ -mixing time $T(\epsilon)$. We have $\lambda(M^{T(\epsilon)}) \leq \sqrt{2\epsilon}$.*

Claim 4.3.2. *Let M be an ergodic Markov chain with state space $[n]$, stationary distribution π , and spectral expansion $\lambda = \lambda(M)$. Let (V_1, \dots, V_t) denote a t -step random walk on M starting from an initial distribution φ on $[n]$, i.e., $V_1 \leftarrow \varphi$. For every $i \in [t]$, let $f_i : [n] \rightarrow [0, 1]$ be a weight function at step i such that the expected weight $\mathbb{E}_{v \leftarrow \pi}[f_i(v)] = \mu$ for all i . Define the total weight of the walk (V_1, \dots, V_t) by $X \triangleq \sum_{i=1}^t f_i(V_i)$. There exists some constant c and a parameter $r > 0$ that depends only on λ and δ such that*

$$\begin{aligned} 1. \quad \frac{\mathbb{E}[e^{rX}]}{e^{r(1+\delta)\mu t}} &\leq \begin{cases} c\|\varphi\|_\pi \exp(-\delta^2(1-\lambda)\mu t/36) & \text{for } 0 \leq \delta \leq 1 \\ c\|\varphi\|_\pi \exp(-\delta(1-\lambda)\mu t/36) & \text{for } \delta > 1. \end{cases} \\ 2. \quad \frac{\mathbb{E}[e^{-rX}]}{e^{-r(1-\delta)\mu t}} &\leq c\|\varphi\|_\pi \exp(-\delta^2(1-\lambda)\mu t/36) \quad \text{for } 0 \leq \delta \leq 1. \end{aligned}$$

Claim 4.3.1 gives a bound on the spectral expansion of each sub-walk $X^{(i)}$, utilizing the fact that they have unit mixing times. Claim 4.3.2 is a spectral version of Chernoff bounds for Markov chains. As stated previously, while similar results exist, we provide our own elementary proof of claim 4.3.2, both for completeness and because it may be of independent interest.

We now continue the proof assuming these two claims. Using Claim 4.3.1, we know $\lambda(N) \leq \frac{1}{2}$. Next, by Claim 4.3.2, for the i -th sub-walk, we have

$$\frac{\mathbb{E}[e^{rX^{(i)}}]}{e^{r(1+\delta)\mu t/T}} \leq \begin{cases} c\|\varphi M^i\|_\pi \exp(-\delta^2\mu t/(72T)) & \text{for } 0 \leq \delta \leq 1 \\ c\|\varphi M^i\|_\pi \exp(-\delta\mu t/(72T)) & \text{for } \delta > 1 \end{cases} \quad (4.5)$$

for an appropriately chosen r (which depends only on λ and δ and hence the same for all i). Note that M^i arises because $X^{(i)}$ starts from the distribution φM^i . On the other hand, notice that $\|\varphi M^i\|_\pi^2 = \|\varphi^\parallel M^i\|_\pi^2 + \|\varphi^\perp M^i\|_\pi^2 \leq \|\varphi^\parallel\|_\pi^2 + \lambda^2(M^i)\|\varphi^\perp\|_\pi^2 \leq \|\varphi\|_\pi^2$ (by using Lemma 4.3.3), or in other words $\|\varphi M^i\|_\pi \leq \|\varphi\|_\pi$. Together with (4.3) and

(4.4), we obtain

$$\Pr[X \geq (1 + \delta)\mu t] \leq \begin{cases} c\|\varphi\|_\pi \exp(-\delta^2 \mu t / (72T)) & \text{for } 0 \leq \delta \leq 1 \\ c\|\varphi\|_\pi \exp(-\delta \mu t / (72T)) & \text{for } \delta > 1 \end{cases}$$

This proves the first half of the theorem. The second case can be proved in a similar manner, namely that

$$\begin{aligned} \Pr[X \leq (1 - \delta)\mu t] &= \Pr\left[\bar{X} \leq \frac{(1 - \delta)\mu t}{T}\right] \\ &\leq \frac{\mathbb{E}[e^{-r\bar{X}}]}{e^{-r(1-\delta)\mu t/T}} \\ &\leq \sum_{k=1}^T \frac{1}{T} \frac{\mathbb{E}[e^{-rX^{(k)}}]}{e^{-r(1-\delta)\mu t/T}} \\ &\leq c\|\varphi\|_\pi \exp(-\delta^2 \mu t / (72T)) \end{aligned}$$

again by Jensen's inequality applied to $\exp(\cdot)$.

□

4.3.1 Mixing Time v.s. Spectral Expansion

We next prove Claim 4.3.1. We remark that Sinclair [100] presents a similar result for reversible Markov chains: for every parameter $\varepsilon \in (0, 1)$,

$$\frac{1}{2} \frac{\lambda(M)}{1 - \lambda(M)} \log \frac{1}{2\varepsilon} \leq T(\varepsilon), \quad (4.6)$$

where $T(\varepsilon)$ is the ε -mixing-time of M . However, in general it is impossible to get a bound on $\lambda(M)$ based on mixing time information for general irreversible chains because a chain M can have $\lambda(M) = 1$ but the ε -mixing-time of M is, say, $T(\varepsilon) = 2$ for some constant ε (and $\lambda(M^2) \ll 1$).

In light of this issue, our proof of Claim 4.3.1 depends crucially on the fact that $M^{T(\varepsilon)}$ has mixing time 1, which, as we shall see, translates to a bound on its spectral expansion that holds regardless of reversibility. We need the following result on reversible Markov chains, which is stronger result than Eq. (4.6) from [100].

Lemma 4.3.2. *Let $0 < \varepsilon \leq 1/2$ be a parameter. Let M be an ergodic reversible Markov chain with ε -mixing time $T(\varepsilon)$ and spectral expansion $\lambda(M)$. It holds that $\lambda(M) \leq (2\varepsilon)^{1/T(\varepsilon)}$.*

We remark that it appears possible to prove Lemma 4.3.2 by adopting an analysis similar to Aldous' [2], who addressed the continuous time case. We present an alternative proof that is arguably simpler; in particular, our proof does not use the spectral representation theorem as used in [2] and does not involve arguments that take the number of steps to infinity.

Proof. (of Lemma 4.3.2) Recall that for an ergodic reversible Markov chain M , it holds that $\lambda(M^t) = \lambda^t(M)$ for every $t \in \mathbb{N}$. Hence, it suffices to show that $\lambda(M^{T(\varepsilon)}) \leq 2\varepsilon$. Also, recall that $\lambda(M^{T(\varepsilon)})$ is simply the second largest eigenvalue (in absolute value) of $M^{T(\varepsilon)}$. Let v be the corresponding eigenvector, i.e. v satisfies $vM^{T(\varepsilon)} = \lambda(M^{T(\varepsilon)})v$. Since M is reversible, the entries of v are real-valued. Also, notice that v is a left eigenvector of M while $(1, 1, \dots, 1)^T$ is a right eigenvector of M (using the fact that each row of M sums to one). Furthermore, v and $(1, \dots, 1)^T$ do not share the same eigenvalue. So we have $\langle v, (1, \dots, 1)^T \rangle = 0$, i.e. $\sum_i v_i = 0$. Therefore, by scaling v , we can assume w.l.o.g. that $x \triangleq v + \pi$ is a distribution. We have the following claim.

Claim 4.3.3. *Let x be an arbitrary initial distribution. Let M be an ergodic Markov*

chain with stationary distribution π and mixing time $T(\epsilon)$. We have $\|xM^{T(\epsilon)} - \pi\|_{TV} \leq 2\epsilon\|x - \pi\|_{TV}$.

Proof. (of Claim 4.3.3) The inequality holds trivially when $x = \pi$. Let $x \neq \pi$ be an arbitrary distribution on M , $\delta \triangleq \|x - \pi\|_{TV} > 0$, and $y \triangleq x - \pi$. We decompose y into a positive component y^+ and a negative component y^- by

$$y_i^+ = \begin{cases} y_i & \text{if } y_i \geq 0 \\ 0 & \text{o.w.} \end{cases} \quad \text{and} \quad y_i^- = \begin{cases} 0 & \text{if } y_i \geq 0 \\ -y_i & \text{o.w.} \end{cases}$$

Note that by definition, $\sum_i y_i^+ = \sum_i y_i^- = \delta$. We define $z^+ = y^+/\delta$ and $z^- = y^-/\delta$. Observe that z^+ and z^- are distributions. By the definition of ϵ -mixing time, we have

$$\|z^+ M^{T(\epsilon)} - \pi\|_{TV} \leq \epsilon, \quad \text{and} \quad \|z^- M^{T(\epsilon)} - \pi\|_{TV} \leq \epsilon,$$

or equivalently, $\|z^+ M^{T(\epsilon)} - \pi\|_1 \leq 2\epsilon$ and $\|z^- M^{T(\epsilon)} - \pi\|_1 \leq 2\epsilon$. Now, we are ready to bound the statistical distance $\|xM^{T(\epsilon)} - \pi\|_{TV}$ as follows.

$$\begin{aligned} \|xM^{T(\epsilon)} - \pi\|_{TV} &= (1/2)\|xM^{T(\epsilon)} - \pi\|_1 \\ &= (1/2)\|(x - \pi)M^{T(\epsilon)}\|_1 \\ &= (1/2)\|(y^+ - y^-)M^{T(\epsilon)}\|_1 \\ &= (1/2)\|\delta z^+ M^{T(\epsilon)} - \delta z^- M^{T(\epsilon)}\|_1 \\ &= (\delta/2)\|(z^+ M^{T(\epsilon)} - \pi) - (z^- M^{T(\epsilon)} - \pi)\|_1 \\ &\leq (\delta/2)(\|z^+ M^{T(\epsilon)} - \pi\|_1 + \|z^- M^{T(\epsilon)} - \pi\|_1) \leq 2\epsilon\delta. \end{aligned}$$

□

We now continue to prove Lemma 4.3.2. By Claim 4.3.3, $\|xM^{T(\epsilon)} - \pi\|_{TV} \leq 2\epsilon\|x - \pi\|_{TV}$, i.e. $\|xM^{T(\epsilon)} - \pi\|_1 \leq 2\epsilon\|x - \pi\|_1$. Observing that $(xM^{T(\epsilon)} - \pi)$ and $(x - \pi)$

are simply $\lambda(M^{T(\varepsilon)})v$ and v , the above inequality means $\lambda(M^{T(\varepsilon)})\|v\|_1 \leq 2\varepsilon\|v\|_1$, which implies $\lambda(M^{T(\varepsilon)}) \leq 2\varepsilon$, as desired. \square

We are now ready to prove our main claim.

Proof. (of Claim 4.3.1) The idea is to reduce to the reversible case by considering the reversiblization of $M^{T(\varepsilon)}$. Let $\tilde{M}^{T(\varepsilon)}$ be the time reversal of $M^{T(\varepsilon)}$, and $R \triangleq M^{T(\varepsilon)}\tilde{M}^{T(\varepsilon)}$ be the reversiblization of $M^{T(\varepsilon)}$. By Claim 4.3.1, $\lambda(M^{T(\varepsilon)}) = \sqrt{\lambda(R)}$. Let us recall (from Chapter 4.2) that M , $M^{T(\varepsilon)}$, and $\tilde{M}^{T(\varepsilon)}$ all share the same stationary distribution π . Next, we claim that the ε -mixing-time of R is 1. This is because $\|\varphi M^{T(\varepsilon)}\tilde{M}^{T(\varepsilon)} - \pi\|_{TV} \leq \|\varphi M^{T(\varepsilon)} - \pi\|_{TV} \leq \varepsilon$, where the second inequality uses the definition of $T(\varepsilon)$ and the first inequality holds since any Markov transition is a contraction mapping: for any Markov transition, say $S = (s(i, j))$, and any vector x , $\|xS\|_1 = \sum_j |\sum_i x_i s(i, j)| \leq \sum_j \sum_i |x_i| s(i, j) = \sum_i |x_i| = \|x\|_1$; putting $x = \varphi M^{T(\varepsilon)} - \pi$ and $S = \tilde{M}^{T(\varepsilon)}$ gives the first inequality. Now, by Lemma 4.3.2, $\lambda(R) \leq 2\varepsilon$, and hence $\lambda(M^{T(\varepsilon)}) = \sqrt{\lambda(R)} \leq \sqrt{2\varepsilon}$, as desired. \square

4.3.2 Bounding the Moment Generating Function

We now prove Claim 4.3.2. We focus on the first inequality in the claim; the derivation of the second inequality is similar and is deferred to Chapter 4.5.

Claim 4.3.2 leads directly to a spectral version of the Chernoff bound for Markov chains. Lezaud [79] and Wagner [106] give similar results for the case where f_i are the same for all i . The analysis of [106] in particular can be extended to the case where the functions f_i are different. Here we present an alternative analysis and along the way will discuss the merit of our approach compared to the previous proofs.

Recall that we define $X = \sum_{i=1}^t f_i(V_i)$. We start with the following observation, which has been used previously [55, 79, 106].

$$\mathbb{E}[e^{rX}] = \|\varphi P_1 M P_2 \dots M P_t\|_1, \quad (4.7)$$

where the P_i are diagonal matrices with diagonal entries $(P_i)_{j,j} \triangleq e^{r f_i(j)}$ for $j \in [n]$. One can verify this fact by observing that each walk V_1, \dots, V_t is assigned the corresponding probability in the product of M 's with the appropriate weight $e^{r \sum_i f_i(V_i)}$.

For ease of exposition, let us assume P_i are all the same at this moment. Let $P = P_1 = \dots = P_t$, then (4.7) becomes $\|\varphi(PM)^{t-1}P\|_1 = \langle \varphi(PM)^{t-1}P, \pi \rangle_\pi = \langle \varphi(PM)^t, \pi \rangle_\pi = \|\varphi(PM)^t\|_1$ (see Lemma 4.3.3 below). Up to this point, our analysis is similar to previous work [48, 55, 79, 106]. Now there are two natural possible ways of bounding $\|\varphi(PM)^t\|_1 = \langle \varphi(PM)^t, \pi \rangle_\pi$.

- Approach 1. Bounding the spectral norm of the matrix PM .** In this approach, we observe that $\langle \varphi(PM)^t, \pi \rangle_\pi \leq \|\varphi\|_\pi \|PM\|_\pi^t$ where $\|PM\|_\pi$ is the operator norm of the matrix PM induced by $\|\cdot\|_\pi$ (see, for example, the proof of Theorem 1 in [106]). This method decouples the effect of each PM as well as the initial distribution. When M is reversible, $\|PM\|_\pi$ can be bounded through Kato's spectral perturbation theory [48, 75, 79]. Alternatively, Wagner [106] tackles the variational description of $\|PM\|_\pi$ directly, using only elementary techniques, whose analysis can be generalized to irreversible chains.
- Approach 2. Inductively giving a bound for $x(PM)^i$ for all $i \leq t$.** In this approach, we do not decouple the product $\varphi(PM)^t$. Instead, we trace the change of the vector $\varphi(PM)^i$ for each $i \leq t$. As far as we know, only

Healy [55] adopts this approach and his analysis is restricted to regular graphs, where the stationary distribution is uniform. His analysis also does not require perturbation theory.

Our proof here generalizes the second approach to any ergodic chains by only using elementary methods. We believe this analysis is more straightforward for the following reasons. First, directly tracing the change of the vector $\varphi(PM)^i$ for each step keeps the geometric insight that would otherwise be lost in the decoupling analysis as in [79, 106]. Second, our analysis studies both the reversible and irreversible chains in a unified manner. We *do not* use the reversibilization technique to address the case for irreversible chains. While the reversibilization technique is a powerful tool to translate an irreversible Markov chain problem into a reversible chain problem, this technique operates in a blackbox manner; proofs based on this technique do not enable us to directly measure the effect of the operator PM .

We now continue our analysis by using a framework similar to the one presented by Healy [55]. We remind the reader that we no longer assume P_i 's are the same. Also, recall that $E[e^{rX}] = \|\varphi P_1 M P_2 \dots M P_t\|_1 = \langle \varphi P_1 M P_2 \dots M P_t, \pi \rangle_\pi = \|(\varphi P_1 M P_2 \dots M P_t)^\parallel\|_\pi$. Let us briefly review the strategy from [55].

- First, we observe that an arbitrary vector x in \mathbb{R}^n can be decomposed into its *parallel* component (with respect to π) $x^\parallel = \langle x, \pi \rangle \pi$ and the *perpendicular* component $x^\perp = x - x^\parallel$ in the L_π space. This decomposition helps tracing the difference (in terms of the norm) between each pair of $\varphi P_1 M \dots P_i M$ and $\varphi P_1 M \dots P_{i+1} M$ for $i \leq t$, i.e. two consecutive steps of the random walk. For this purpose, we need to understand the *effects of the linear operators* M and

P_i when they are applied to an arbitrary vector.

- Second, after we compute the difference between each pair $xP_1M\dots P_iM$ and $xP_1M\dots P_{i+1}M$, we set up a *recursive relation*, the solution of which yields the Chernoff bound.

We now follow this step step framework to prove Claim 4.3.2

The effects of the M and P_i operators Our way of tracing the vector $\varphi P_1M P_2\dots M P_t$ relies on the following two lemmas.

Lemma 4.3.3. (The effect of the M operator) *Let M be an ergodic Markov chain with state space $[n]$, stationary distribution π , and spectral expansion $\lambda = \lambda(M)$. Then*

1. $\pi M = \pi$.
2. For every vector y with $y \perp \pi$, we have $yM \perp \pi$ and $\|yM\|_\pi \leq \lambda \|y\|_\pi$.

Lemma 4.3.4. (The effect of the P operator) *Let M be an ergodic Markov chain with state space $[n]$ and stationary distribution π . Let $f : [n] \rightarrow [0, 1]$ be a weight function with $\mathbb{E}_{v \leftarrow \pi}[f(v)] = \mu$. Let P be a diagonal matrix with diagonal entries $P_{j,j} \triangleq e^{rf(j)}$ for $j \in [n]$, where r is a parameter satisfying $0 \leq r \leq 1/2$. Then*

1. $\|(\pi P)^\parallel\|_\pi \leq 1 + (e^r - 1)\mu$.
2. $\|(\pi P)^\perp\|_\pi \leq 2r\sqrt{\mu}$.
3. For every vector $y \perp \pi$, $\|(yP)^\parallel\|_\pi \leq 2r\sqrt{\mu}\|y\|_\pi$.
4. For every vector $y \perp \pi$, $\|(yP)^\perp\|_\pi \leq e^r\|y\|_\pi$.

Items 1 and 4 of Lemma 4.3.4 state that P can stretch both the perpendicular and parallel components along their original directions moderately. Specifically, a parallel vector is stretched by at most a factor of $(1 + (e^r - 1)\mu) \approx 1 + O(r\mu)$ and a perpendicular vector is stretched by a factor of at most $e^r \approx 1 + O(r)$. (Recall r will be small.) On the other hand, items 2 and 3 of the lemma state that P can create a new perpendicular component from a parallel component and vice versa, but the new component is of a much smaller size compared to the original component (i.e. only of length at most $2r\sqrt{\mu}$ times the original component).

Remark We note that the key improvement of our analysis over that of Healy [55] stems from items 2 and 3 of Lemma 4.3.4. Healy [55] proved a bound with a factor of $(e^r - 1)/2 = O(r)$ for both items for the special case of undirected and regular graphs. Our quantitative improvement to $O(r\sqrt{\mu})$ (which is tight) is the key for us to prove a multiplicative Chernoff bound without any restriction on the spectral expansion of M .

Note that Lemma 4.3.3 is immediate from the definitions of π and λ . We focus on the proof of Lemma 4.3.4:

Proof. (of Lemma 4.3.4). For the first item, note that by definition, $\|(\pi P)^\parallel\|_\pi = \langle \pi P, \pi \rangle_\pi = \sum_i e^{rf(i)} \pi_i$. We simplify the sum using the fact that $e^{rx} \leq 1 + (e^r - 1)x$ when $r, x \in [0, 1]$.

$$\|(\pi P)^\parallel\|_\pi = \sum_i e^{rf(i)} \pi_i \leq \sum_i (1 + (e^r - 1)f(i)) \pi_i = \sum_i \pi_i + (e^r - 1) \sum_i f(i) \pi_i = 1 + (e^r - 1)\mu,$$

where the last equality uses the fact that $\sum_i \pi_i = 1$, and $\sum_i f(i) \pi_i = \mathbb{E}_{v \leftarrow \pi}[f(v)] = \mu$.

For the second item, by the Pythagorean theorem, we have

$$\|(\pi P)^\perp\|_\pi^2 = \|\pi P\|_\pi^2 - \|(\pi P)^\parallel\|_\pi^2 = \sum_i e^{2rf(i)}\pi_i - \left(\sum_i e^{rf(i)}\pi_i\right)^2.$$

Recall that $r \leq 1/2$ and $f(i) \leq 1$, and therefore $2rf(i) \leq 1$. Using the fact that $1 + x \leq e^x \leq 1 + x + x^2$ when $x \in [0, 1]$, we have

$$\begin{aligned} & \sum_i e^{2rf(i)}\pi_i - \left(\sum_i e^{rf(i)}\pi_i\right)^2 \\ & \leq \sum_i (1 + 2rf(i) + 4r^2f^2(i))\pi_i - \left(\sum_i (1 + rf(i))\pi_i\right)^2 \\ & \leq 1 + 2r\mu + 4r^2\mu - (1 + r\mu)^2 \\ & = 1 + 2r\mu + 4r^2\mu - (1 + 2r\mu + r^2\mu^2) \leq 4r^2\mu, \end{aligned}$$

The second inequality uses the fact that $\sum_i f^2(i)\pi(i) \leq \sum_i f(i)\pi(i) = \mu$ (since $0 \leq f(i) \leq 1$). It follows that $\|(\pi P)^\perp\|_\pi \leq \sqrt{4r^2\mu} = 2r\sqrt{\mu}$.

For the third item, by definition, $\|(yP)^\parallel\|_\pi = \langle yP, \pi \rangle_\pi$. Since P is diagonal, we have $\langle yP, \pi \rangle_\pi = \langle y, \pi P \rangle_\pi$. By definition, $y \perp \pi$ means $\langle y, \pi \rangle_\pi = 0$. Therefore, $\|(yP)^\parallel\|_\pi = \langle y, \pi P \rangle_\pi - \langle y, \pi \rangle_\pi = \langle y, \pi(P - I) \rangle_\pi$. By the Cauchy-Schwarz Inequality, we have $\langle y, \pi(P - I) \rangle_\pi \leq \|y\|_\pi \|\pi(P - I)\|_\pi$.

We proceed to upper bound $\|\pi(P - I)\|_\pi$:

$$\|\pi(P - I)\|_\pi^2 = \sum_i (\pi_i(e^{rf(i)} - 1))^2 / \pi_i = \sum_i (e^{rf(i)} - 1)^2 \pi_i.$$

Using $e^{rx} \leq 1 + (e^r - 1)x$ for $r, x \in [0, 1]$, we have $\sum_i (e^{rf(i)} - 1)^2 \pi_i \leq \sum_i (1 + (e^r - 1)f(i) - 1)^2 \pi_i = \sum_i (e^r - 1)^2 f^2(i) \pi_i \leq (2r)^2 \sum_i f(i) \pi_i \leq (2r)^2 \mu$, where the second-to-last inequality uses the fact that $e^r - 1 \leq 2r$ for $r \in [0, 1]$ and $0 \leq f(i) \leq 1$. Therefore, $\|(yP)^\parallel\|_\pi \leq \|\pi(P - I)\|_\pi \|y\|_\pi \leq 2r\sqrt{\mu} \|y\|_\pi$.

Finally, for the fourth item, we have

$$\|(yP)^\perp\|_\pi^2 \leq \|yP\|_\pi^2 = \sum_i \frac{y_i^2 e^{2rf(i)}}{\pi_i} \leq \sum_i \frac{y_i^2 e^{2r}}{\pi_i} = e^{2r} \|y\|_\pi^2,$$

which implies $\|(yP)^\perp\| \leq e^r \|y\|_\pi$. \square

Recursive analysis We now provide a recursive analysis for the terms $xP_1M\dots MP_i$ for $i \leq t$ based on our understanding of the effects from the linear operators M and P_i . This completes the proof for Claim 4.3.2.

Proof. (of Claim 4.3.2). First, recall that

$$\mathbb{E}[e^{rX}] = \|(\varphi P_1 M P_2 \dots M P_t)^\parallel\|_\pi = \|(\varphi P_1 M P_2 \dots M P_t M)^\parallel\|_\pi = \left\| \left(\varphi \prod_{i=1}^t (P_i M) \right)^\parallel \right\|_\pi.$$

where the second equality comes from Lemma 4.3.3. Our choice of r is

$$r = \min\{1/2, \log(1/\lambda)/2, 1 - \sqrt{\lambda}, (1 - \lambda)\delta/18\}.$$

We shall explain how we make such a choice as we walk through our analysis.

We now trace the π -norm of both parallel and perpendicular components of the random walk for each application of $P_i M$. Let $z_0 \triangleq \varphi$ and $z_i = z_{i-1} P_i M$ for $i \in [t]$.

By triangle inequality and Lemma 4.3.3 and 4.3.4, for every $i \in [t]$,

$$\begin{aligned} \|z_i^\parallel\|_\pi &= \|(z_{i-1} P_i M)^\parallel\|_\pi = \|((z_{i-1}^\parallel + z_{i-1}^\perp) P_i M)^\parallel\|_\pi \\ &\leq \|z_{i-1}^\parallel P_i M\|_\pi + \|z_{i-1}^\perp P_i M\|_\pi \\ &\leq (1 + (e^r - 1)\mu) \|z_{i-1}^\parallel\|_\pi + (2r\sqrt{\mu}) \|z_{i-1}^\perp\|_\pi, \end{aligned}$$

and similarly,

$$\begin{aligned} \|z_i^\perp\|_\pi &\leq \|(z_{i-1}^\parallel P_i M)^\perp\|_\pi + \|(z_{i-1}^\perp P_i M)^\perp\|_\pi \leq (2r\lambda\sqrt{\mu}) \|z_{i-1}^\parallel\|_\pi + (e^r \lambda) \|z_{i-1}^\perp\|_\pi \\ &\leq (2r\lambda\sqrt{\mu}) \|z_{i-1}^\parallel\|_\pi + \sqrt{\lambda} \|z_{i-1}^\perp\|_\pi, \end{aligned}$$

where the last inequality holds when $r \leq (1/2) \log(1/\lambda)$ i.e. $e^r \leq 1/\sqrt{\lambda}$. The reason to require $r \leq (1/2) \log(1/\lambda)$ is that we can guarantee the perpendicular component is *shrinking* (by a factor of $\sqrt{\lambda} < 1$) after each step.

Now let $\alpha_0 = \|z_0^\parallel\|_\pi = 1$ and $\beta_0 = \|z_0^\perp\|_\pi$, and define for $i \in [t]$,

$$\alpha_i = (1 + (e^r - 1)\mu) \alpha_{i-1} + (2r\sqrt{\mu}) \beta_{i-1} \quad \text{and} \quad \beta_i = (2r\lambda\sqrt{\mu}) \alpha_{i-1} + \sqrt{\lambda} \beta_{i-1}.$$

One can prove by induction easily that $\|z_i^\parallel\|_\pi \leq \alpha_i$ and $\|z_i^\perp\|_\pi \leq \beta_i$ for every $i \in [t]$, and α_i 's are strictly increasing. Therefore, bounding the moment generating function $E[e^{rX}] = \|z_t^\parallel\|_\pi \leq \alpha_t$ boils down to bounding the recurrence relation for α_i and β_i .

Observe that in the recurrence relation, only the coefficient $(1 + (e^r - 1)\mu) > 1$ while the remaining coefficients $(2r\sqrt{\mu})$, $(2r\lambda\sqrt{\mu})$, and $\sqrt{\lambda}$ are all less than 1 if r is chosen sufficiently small. This suggests, intuitively, α_i 's terms will eventually dominate. This provides us a guide to reduce the recurrence relation to a single variable as follows.

First let us give an upper bound for β_i .

Claim 4.3.4. *For every $i \in [t]$, $\beta_i \leq 2r \left(\sum_{j=0}^{i-1} \sqrt{\lambda^{j+2}\mu} \right) \alpha_{i-1} + \sqrt{\lambda^i} \beta_0$.*

Proof. of Claim 4.3.4. The lemma follows by expanding the recurrence relation and

using the fact that α_i 's are increasing. i.e.

$$\begin{aligned}\beta_i &= 2r\lambda\sqrt{\mu_i}\alpha_{i-1} + \sqrt{\lambda}\beta_{i-1} = 2r\lambda\sqrt{\mu}\alpha_{i-1} + \sqrt{\lambda}2r\lambda\sqrt{\mu}\alpha_{i-2} + \sqrt{\lambda^2}\beta_{i-2} \\ &= \dots = 2r\left(\sum_{j=0}^{i-1}\sqrt{\lambda^{j+2}\mu}\alpha_{i-j-1}\right) + \sqrt{\lambda^i}\beta_0\end{aligned}$$

Finally, by using the fact that α_i are strictly increasing, we complete the proof. \square

We can then bound α_i by substituting β_{i-1} using Claim 4.3.4.

Claim 4.3.5. $\alpha_1 \leq (1 + (e^r - 1)\mu) + 2r\sqrt{\mu}\beta_0$, and for every $2 \leq i \leq t$,

$$\alpha_i \leq \left(1 + (e^r - 1)\mu + 4r^2\sqrt{\mu}\left(\sum_{j=0}^{i-2}\sqrt{\lambda^{j+2}\mu}\right)\right)\alpha_{i-1} + 2r\sqrt{\lambda^{i-1}\mu}\beta_0.$$

Proof. The case of $i = 1$ is trivial. For $2 \leq i \leq t$, this follows by applying the recurrence relation, Claim 4.3.4, and the fact that $\alpha_{i-2} < \alpha_{i-1}$.

$$\begin{aligned}\alpha_i &= (1 + (e^r - 1)\mu)\alpha_{i-1} + (2r\sqrt{\mu})\beta_{i-1} \\ &\leq (1 + (e^r - 1)\mu)\alpha_{i-1} + (2r\sqrt{\mu})\left(2r\left(\sum_{j=0}^{i-2}\sqrt{\lambda^{j+2}\mu}\right)\alpha_{i-2} + \sqrt{\lambda^{i-1}}\beta_0\right) \\ &\leq \left(1 + (e^r - 1)\mu + 4r^2\sqrt{\mu}\left(\sum_{j=0}^{i-2}\sqrt{\lambda^{j+2}\mu}\right)\right)\alpha_{i-1} + 2r\sqrt{\lambda^{i-1}\mu}\beta_0\end{aligned}$$

\square

For notational simplicity, let $A_1 = 1 + (e^r - 1)$ and for $1 < i \leq t$, let

$$A_i \triangleq \left(1 + (e^r - 1)\mu + 4r^2\sqrt{\mu}\left(\sum_{j=0}^{i-2}\sqrt{\lambda^{j+2}\mu}\right)\right).$$

Claim 4.3.5 then can be expressed as $\alpha_i \leq A_i\alpha_{i-1} + 2r\sqrt{\mu}\min\{\sqrt{\lambda^{i-1}}, 1\}\beta_0$, for every $i \in [t]$. By expanding iteratively, we obtain (4.8), where the last inequality uses the fact that $1/(1 - \sqrt{\lambda}) \leq 2/(1 - \lambda)$ for $\lambda \in [0, 1)$. It remains to upper bound $\prod_i A_i$.

$$\begin{aligned}
& \alpha_t \\
& \leq A_t(A_{t-1}(\cdots(A_3(A_2(A_1 + 2r\sqrt{\mu}\beta_0) + 2r\sqrt{\lambda\mu}\beta_0) + 2r\sqrt{\lambda^2\mu}\beta_0)\cdots) + 2r\sqrt{\lambda^{t-2}\mu}\beta_0) + 2r\sqrt{\lambda^{t-1}\mu}\beta_0 \\
& = (A_t \cdots A_1) + (A_t \cdots A_2(2r\sqrt{\mu}\beta_0)) + (A_t \cdots A_3(2r\sqrt{\lambda\mu}\beta_0)) + \cdots + A_t(2r\sqrt{\lambda^{t-2}\mu}\beta_0) + 2r\sqrt{\lambda^{t-1}\mu}\beta_0 \\
& \leq \left(1 + 2r\sqrt{\mu}\beta_0 + 2r\sqrt{\lambda\mu}\beta_0 + 2r\sqrt{\lambda^2\mu}\beta_0 + \cdots + 2r\sqrt{\lambda^{t-1}\mu}\beta_0\right) (\prod_i A_i) \\
& \leq \left(1 + \frac{4r\sqrt{\mu}\beta_0}{1-\sqrt{\lambda}}\right) (\prod_i A_i) \leq \left(1 + \frac{8r\sqrt{\mu}\beta_0}{1-\lambda}\right) (\prod_i A_i),
\end{aligned} \tag{4.8}$$

Using $(1 + x) \leq e^x$, we have

$$\prod_{i=1}^t A_i \leq \exp \left\{ (e^r - 1)\mu + \sum_{i=2}^t \left((e^r - 1)\mu_i + 4r^2 \sqrt{\mu} \left(\sum_{j=0}^{i-2} \sqrt{\lambda^{j+2}\mu} \right) \right) \right\}.$$

The first two sums in the exponent lead to $\sum_i (e^r - 1)\mu_i = (e^r - 1)\mu t$. We now bound the last sum in the exponent, which can be viewed as an “error” term due to the correlation between each step of the random walk.

$$\sum_{i=2}^t 4r^2 \sqrt{\mu} \sum_{j=0}^{i-2} \sqrt{\lambda^{j+2}\mu} \leq 4r^2 \mu \sum_{i=1}^t \sum_{j=0}^{i-2} \sqrt{\lambda^j} = 4r^2 \mu t \sum_{j=0}^{t-2} \sqrt{\lambda^j} \leq \frac{8r^2 \mu t}{1 - \lambda},$$

where last inequality uses $\sum_{j=0}^{t-2} \sqrt{\lambda^j} \leq 1/(1 - \sqrt{\lambda}) \leq 2/(1 - \lambda)$. Putting things together, we have

$$\prod_{i=1}^t A_i \leq \exp \left\{ (e^r - 1)\mu t + \frac{8r^2 \mu t}{1 - \lambda} \right\} = \exp \left\{ \left((e^r - 1) + \frac{8r^2}{1 - \lambda} \right) \mu t \right\},$$

and recalling that $\|\varphi\|_\pi = 1$ and $\beta_0 = \|\varphi^\perp\|_\pi$,

$$\mathbb{E}[e^{rX}] \leq \alpha_t \leq \left(1 + \frac{8r\mu\beta_0}{1 - \lambda} \right) \left(\prod_i A_i \right) \leq 2 \max \left\{ 1, \frac{8r\sqrt{\mu}}{1 - \lambda} \right\} \|\varphi\|_\pi \exp \left\{ \left((e^r - 1) + \frac{8r^2}{(1 - \lambda)} \right) \mu t \right\}.$$

Recall that our goal is to choose an r to bound $\mathbb{E}[e^{rX}]/e^{r(1+\delta)\mu t}$. Choosing $r = \min\{1/2, \log(1/\lambda)/2, 1 - \sqrt{\lambda}, (1 - \lambda)\delta/18\} = (1 - \lambda)\delta/18$, we complete the proof of Claim 4.3.2. \square

Notice that our proof also works even for the case $\mathbb{E}_\pi[f_i(v)]$ are *different* for different values of i , which results in a more general Chernoff type bound based on spectral expansions. This more general result, as far as we know, has not been noted in existing literatures with the exception of Healy [55], who gave a Chernoff bound of this kind with stronger assumptions for regular graphs, although the analysis given

by Lezaud [79] or Wagner [106] also appears to be generalizable as well. On the other hand, this strengthened result of Claim 4.3.2 does not seem to be sufficient to remove the requirement that $E_\pi[f_i(v)]$ are the same for Theorem 4.3.1.

4.3.3 Continuous Time Case

We now generalize our main result to cover the continuous time chains. The analysis is similar to the one presented by Lezaud [79].

Theorem 4.3.5. *Let Λ be the generator of an ergodic continuous time Markov chain with state space $[n]$ and mixing time $T = T(\epsilon)$. Let $\{v_t : t \in \mathbf{R}^+\}$ be a random walk on the chain starting from an initial distribution φ such that v_t represents the state where the walk stay at time t . Let $\{f_t : [n] \rightarrow [0, 1] \mid t \in \mathbf{R}^+\}$ be a family of functions such that $\mu = E_{v \leftarrow \pi}[f_t(v)]$ for all t . Define the weight over the walk $\{v_s : s \in \mathbf{R}^+\}$ up to time t by $X_t \triangleq \int_0^t f_s(v_s)ds$. There exists a constant c such that*

$$\begin{aligned} 1. \Pr[X \geq (1 + \delta)\mu t] &\leq \begin{cases} c\|\varphi\|_\pi \exp(-\delta^2\mu t/(72T)) & \text{for } 0 \leq \delta \leq 1 \\ c\|\varphi\|_\pi \exp(-\delta\mu t/(72T)) & \text{for } \delta > 1 \end{cases} \\ 2. \Pr[X \leq (1 - \delta)\mu t] &\leq c\|\varphi\|_\pi \exp(-\delta^2\mu t/(72T)) \quad \text{for } 0 \leq \delta \leq 1 \end{aligned}$$

Proof. (of Theorem 4.3.5). We mimic the strategy from Lezaud [79] to discretize the chain in b time units, i.e. consider the states v_{ib} for $i = 0, 1, \dots, t/b$. The stationary distribution of this discretized chain v_{ib} is the same as the original continuous time

chain, and hence $\mu = E_\pi f_t(v_t) = E_\pi f_{ib}(v_{ib})$. Now by Theorem 4.3.1 we have

$$\begin{aligned} 1. \Pr \left[\sum_{i=1}^{t/b} f_{ib}(v_{ib}) \geq (1 + \delta) \frac{\delta \mu t}{b} \right] &\leq \begin{cases} c \|\varphi\|_\pi \exp(-\delta^2 \mu(t/b)/(72T/b)) & \text{for } 0 \leq \delta \leq 1 \\ c \|\varphi\|_\pi \exp(-\delta \mu(t/b)/(72T/b)) & \text{for } \delta > 1 \end{cases} \\ 2. \Pr \left[\sum_{i=1}^{t/b} f_{ib}(v_{ib}) \leq (1 - \delta) \frac{\delta \mu t}{b} \right] &\leq c \|\varphi\|_\pi \exp(-\delta^2 \mu(t/b)/(72T/b)) \quad \text{for } 0 \leq \delta \leq 1 \end{aligned}$$

Notice that the mixing time for the discretized chain is T/b while the total number of steps here is t/b . In the exponents, the term b appears in both the numerator and the denominator and they cancel with each other. Taking limit as $b \rightarrow 0$ completes the proof [79]. \square

4.4 Construction of Mixing Markov Chain with No Spectral Expansion

We now show that any ergodic Markov chain M with mixing time $T = T(1/4)$ can be modified to a chain M' such that M' has mixing time $O(T)$ but spectral expansion $\lambda(M') = 1$.

Our modification is based on the following simple observation. Let M' be an ergodic Markov chain with stationary distribution π' . If there exist two states v and v' such that (i) $M'_{v,v'} = 1$, i.e., state v leaves to state v' with probability 1, and (ii) $M'_{u,v'} = 0$ for all $u \neq v$, i.e., the only state transits to v' is v , then $\lambda(M') = 1$: Note that in this case, $\pi'(v) = \pi'(v')$ since all probability mass from v leaves to v' , which receives probability mass only from v . Consider a distribution x whose probability mass all concentrates at v , i.e., $x_v = 1$ and $x_u = 0$ for all $u \neq v$. One step walk from

x results in the distribution xM' whose probability mass all concentrates at v' . By definition, $\|x\|_{\pi'} = \|xM'\|_{\pi'}$ and thus $\lambda(M') = 1$.

Now, let M be an ergodic Markov chain with mixing time $T = T(1/4)$ and stationary distribution π . We shall modify M to a Markov chain M' that preserves the mixing-time and satisfies the above property. We mention that it is not hard to modify M to satisfy the above property. The challenge is to do so while preserving the mixing-time. Our construction is as follows.

- For every state v in M , we “split” it into three states $(v, in), (v, mid), (v, out)$ in M' .
- For every state (v, in) in M' , we set $M'_{(v, in), (v, in)} = M'_{(v, in), (v, mid)} = 1/2$, i.e., (v, in) stays in the same state with probability $1/2$ and transits to (v, mid) with probability $1/2$.
- For every state (v, mid) in M' , we set $M'_{(v, mid), (v, out)} = 1$, i.e., (v, mid) always leaves to (v, out) .
- For every pairs of states u, v in M , we set the transition probability $M'_{(u, out), (v, in)}$ from (u, out) to (v, in) to be $M_{u, v}$.

It is not hard to verify that the modified chain M' is well-defined, ergodic, and satisfies the aforementioned property (namely, (v, mid) leaves to (v, out) with probability 1 and is the only state that transits to (v, out)). It remains to show that M' has mixing-time $O(T)$. Toward this goal, let us define yet another Markov chain C that consists of three states $\{in, mid, out\}$ with transition probability $C_{in, in} = C_{in, mid} = 1/2$, and $C_{mid, out} = C_{out, in} = 1$. Clearly, C is ergodic and has constant mixing-time.

Now, the key observation is that a random walk on M' can be decomposed into walks on M and C in the following sense: every step on M' corresponding to a step on C in a natural way, and one step on M' from (u, out) to (v, in) can be identified as a step from u to v in M . Note that the walks on M and C are independent, and in expectation, every 4 steps of walk on M' induce one step of walk on M . It is not hard to see from these observation that the mixing time of M' is at most $8T$.

4.5 The Bound When the Sum Is Less Than Mean

We now prove the remaining part of Claim 4.3.2 for completeness, i.e.

Claim 4.5.1. *Let M be an ergodic Markov chain with state space $[n]$, stationary distribution π , and spectral expansion $\lambda = \lambda(M)$. Let (V_1, \dots, V_t) denote a t -step random walk on M starting from an initial distribution φ on $[n]$, i.e., $V_1 \leftarrow \varphi$. For every $i \in [t]$, let $f_i : [n] \rightarrow [0, 1]$ be a weight function at step i such that the expected weight $\mathbb{E}_{v \leftarrow \pi}[f_i(v)] = \mu$ for all i . Define the total weight of the walk (V_1, \dots, V_t) by $X \triangleq \sum_{i=1}^t f_i(V_i)$. There exists some constant c and a parameter $r > 0$ that depends only on λ and δ such that*

$$2. \frac{\mathbb{E}[e^{-rX}]}{e^{-r(1-\delta)\mu t}} \leq c \|\varphi\|_\pi \exp(-\delta^2(1-\lambda)\mu t/36) \quad \text{for } 0 \leq \delta \leq 1.$$

We mimic the proof strategy presented in Chapter 4.3.2. Observe first that

$$\mathbb{E}[e^{-rX}] = \|xP_1MP_2\dots MP_t\|_1,$$

where P_i 's are diagonal matrices with diagonal entries $(P_i)_{j,j} \triangleq e^{-rf_i(j)}$ for $j \in [n]$.

Thus, our goal is to bound the moment generating function $\mathbb{E}[e^{rX}]$.

Similar to the analysis presented in Chapter 4.3.2, we need to understand the effect of the P_i operators.

Lemma 4.5.1. *Let M be an ergodic Markov chain with state space $[n]$ and stationary distribution π . Let $f : [n] \rightarrow [0, 1]$ be a weight function with $\mathbb{E}_{v \leftarrow \pi}[f(v)] = \mu$. Let P be a diagonal matrix with diagonal entries $P_{j,j} \triangleq e^{-rf(j)}$ for $j \in [n]$, where r is a parameter satisfying $0 \leq r \leq 1/2$. We have*

- $\|(\pi P)^\parallel\|_\pi \leq 1 - r\mu + \frac{r^2}{2}\mu.$
- $\|(\pi P)^\perp\|_\pi \leq \sqrt{2}r\sqrt{\mu}$
- For every vector $y \perp \pi$, $\|(yP)^\parallel\|_\pi \leq r\sqrt{\mu}\|y\|_\pi.$
- For every vector $y \perp \pi$, $\|(yP)^\perp\|_\pi \leq \|y\|_\pi$

Proof. For the first item, we have

$$\begin{aligned}
 \|(\pi P)^\parallel\|_\pi &= \sum_{i \leq n} e^{-rf(i)} \pi_i \\
 &\leq \sum_{i \leq n} (1 - rf(i) + \frac{r^2}{2}f(i)) \pi_i \\
 &\leq 1 - r\mu + \frac{r^2}{2}\mu
 \end{aligned}$$

The first inequality holds because $e^{-rx} \leq 1 - rx + r^2x/2$ for $0 \leq x \leq 1$.

(2). we may use Pythagorean theorem and get

$$\begin{aligned}
\|(\pi P)^\perp\|_\pi^2 &= \|(\pi P)\|_\pi^2 - \|(\pi P)^\parallel\|_\pi^2 \\
&= \sum_{i \leq n} e^{-2rf(i)} \pi_i - \left(\sum_{i \leq n} e^{-rf(i)} \pi_i \right)^2 \\
&\leq \sum_{i \leq n} (1 - 2rf(i) + 2r^2 f^2(i)) \pi_i - \left(\sum_{i \leq n} (1 - rf(i)) \pi_i \right)^2 \\
&= 2r^2 \mu - r^2 \mu^2 \\
&\leq 2r^2 \mu.
\end{aligned}$$

This implies $\|(\pi P)^\perp\|_\pi \leq \sqrt{2}r\sqrt{\mu}$.

(3). First, since $y \perp \pi$, we have $\langle y, \pi \rangle_\pi = 0$. Next notice that by Cauchy Schwarz inequality,

$$\|(yP)^\parallel\|_\pi = \langle y, \pi P \rangle_\pi - \langle y, \pi I \rangle_\pi = \langle y, \pi(P - I) \rangle_\pi \leq \|y\|_\pi \|\pi(P - I)\|_\pi.$$

We next bound $\|\pi(P - I)\|_\pi$. Specifically,

$$\begin{aligned}
\|\pi(P - I)\|_\pi^2 &= \sum_{i \leq n} (e^{-rf(i)} - 1)^2 \pi_i \\
&= \sum_{i \leq n} (1 - e^{-rf(i)})^2 \pi_i \\
&\leq \sum_{i \leq n} (rf(i))^2 \pi_i \\
&\leq r^2 \sum_{i \leq n} f(i) \pi_i \\
&\leq r^2 \mu.
\end{aligned}$$

Therefore, $\|(yP)^\parallel\|_\pi \leq r\sqrt{\mu}\|y\|_\pi$.

(4). We have $\|(yP)^\perp\|_\pi \leq \|(yP)\|_\pi \leq \|y\|_\pi$. □

Now we proceed to prove Claim 4.5.1 using Lemma 4.5.1.

Proof. (of Claim 4.5.1). Let us recall that $z_0 \triangleq x$ and $z_i = z_{i-1}P_iM$ for $i \in [t]$.

Lemma 4.5.1 gives us

$$\|z_i^\parallel\|_\pi \leq (1 - r\mu + \frac{r^2}{2}\mu)\|z_{i-1}^\parallel\|_\pi + r\sqrt{\mu}\|z_{i-1}^\perp\|_\pi$$

and

$$\|z_i^\perp\|_\pi \leq \sqrt{2}\lambda r\sqrt{\mu}\|z_{i-1}^\parallel\|_\pi + \lambda\|z_{i-1}^\perp\|_\pi$$

Following our strategy presented in Chapter 4.3.2, let $\alpha_0 = \|z_0^\parallel\|_\pi = 1$ and $\beta_0 = \|z_0^\perp\|_\pi$ and define for each $i \in [t]$,

$$\alpha_i = (1 - r\mu + \mu r^2/2)\alpha_{i-1} + r\sqrt{\mu}\beta_{i-1} \quad (4.9)$$

and

$$\beta_i = (\sqrt{2}r\lambda\sqrt{\mu})\alpha_{i-1} + \lambda\beta_{i-1}. \quad (4.10)$$

We can inductively show that $\|z_i^\parallel\|_\pi \leq \alpha_i$ and $\|z_i^\perp\|_\pi \leq \beta_i$ for each $i \in [t]$.

Our goal becomes to give an upper bound for α_i and β_i . Also, we shall set $r = \min\{1/2, \log(1/\lambda)/2, 1 - \sqrt{\lambda}, (1 - \lambda)\delta/8\}$ throughout our analysis. Next, we recursively substitute the value of β_i from Eq.(4.10) into Eq.(4.9) and yield,

$$\alpha_i = (1 - (r - r^2/2)\mu)\alpha_{i-1} + \sqrt{2}r^2\mu\lambda\alpha_{i-2} + \dots + \sqrt{2}r^2\mu\lambda^{i-1}\alpha_0 + r\sqrt{\mu}\lambda^{i-1}\beta_0 \quad (4.11)$$

Using the fact that $r \leq 1 - \sqrt{\lambda}$ and thus $\alpha_i \leq (1 - (r - r^2/2)\mu)\alpha_{i-1}$ for all $i \geq 1$, we may conclude $\sqrt{\lambda}\alpha_{i-1} \leq \alpha_i$. Now (4.11) becomes

$$\alpha \leq \left(1 - (r - r^2/2)\mu + \sqrt{2}r^2\sqrt{\mu} \left(\sum_{j=1}^{i-1} \sqrt{\mu}\sqrt{\lambda^{i-j}}\right)\right) \alpha_{i-1} + r\sqrt{\mu}\lambda^{i-1}\beta_0. \quad (4.12)$$

Next, let us define A_i as follows,

$$A_i \triangleq \left(1 - (r - r^2/2)\mu + \sqrt{2}r^2\sqrt{\mu} \left(\sum_{j=0}^{i-1} \sqrt{\mu}\sqrt{\lambda^{i-j}}\right)\right).$$

We then have

$$\alpha_i \leq A_i \alpha_{i-1} + r\sqrt{\mu}\lambda^{i-1}\beta_0.$$

Therefore, we can see that

$$\alpha_t \leq \left(\prod_{i \leq t} A_i \right) \left(1 + \beta_0 \frac{r\sqrt{\mu}}{1-\lambda} \right).$$

On the other hand, we can see that

$$\begin{aligned} \left(\prod_{1 \leq i \leq t} A_i \right) &\leq \exp \left\{ \sum_{i \leq t} (-(r - r^2/2)\mu) + \sum_{1 \leq i \leq t} \sqrt{2}r^2\sqrt{\mu} \left(\sum_{1 \leq j \leq t-1} \sqrt{\lambda^{i-j}\mu} \right) \right\} \\ &\leq \exp \left\{ -(r - r^2/2)\mu t + \frac{2\sqrt{2}r^2}{1-\lambda}\mu t \right\} \\ &= \exp \left\{ -r\mu t + \left(\frac{r^2}{2} + \frac{2\sqrt{2}r^2}{1-\lambda} \right) \mu t \right\} \\ &\leq \exp \left\{ -r\mu t + \left(\frac{4r^2}{1-\lambda} \mu t \right) \right\} \end{aligned}$$

Notice that $1 + \beta_0 \frac{r\sqrt{\mu}}{1-\lambda} = O\left(\frac{r\|x\|_\pi}{1-\lambda}\right)$. By using the fact $r = \min\{1/2, \log(1/\lambda)/2, 1 - \sqrt{\lambda}, (1-\lambda)\delta/8\}$, we complete the proof. \square

Chapter 5

Technology diffusion in communication networks

5.1 Introduction

There has been significant interest in the networking community on the impact of cascade effects on the diffusion of technology upgrades in the Internet [10, 19, 24, 38, 39, 47, 53, 65, 66, 89]. Thinking of the global Internet as a graph, where each node represents an independent, economically-motivated *autonomous system* (AS), *e.g.* AT&T, Google, Telecom Italia, or Bank of America, a key problem is to determine the set of nodes that governments and regulatory groups should target as early adopters of the new technology, with the goal of triggering a cascade that causes more and more nodes to voluntarily adopt the new technology [19, 47, 53, 66]. Given the effort and expense required to target ASes as early adopters, a natural objective (that has appeared in both the networking literature [10, 19, 47] and also that of viral

marketing [37,69]) is to find the smallest possible *seedset* of early adopters that could drive a cascade of adoption; doing this would shed light on how best to manage the upgrade from insecure routing [18] to secure routing [70,76], or from IPv4 to IPv6 [33], or the deployment of technology upgrades like QoS [60], fault localization [12], and denial of service prevention [111].

Thus far, the literature has offered only heuristic solutions this problem. In this paper, we design the first approximation algorithm with a provable performance guarantee that optimizes the selection of early adopter nodes, in a model of that captures the following important property: the technologies we study only allow a pair of nodes to communicate if they have a *path* between them consisting of nodes that also use the new technology [12,19,47,60,63,76,111].

Model. Consider a graph $G(V, E)$ that represents the internetwork. We use the following progressive process to model the diffusion of a new technology: a node starts out as inactive (using an older version of the technology) and *activates* (adopts the new, improved technology) once it obtains sufficient utility from the new technology. Once a node is active, it can never become inactive. To model the cost of technology deployment, the standard approach [50,69,97] is to associate a threshold $\theta(u)$ with each node u that determines how large its utility should be before it is willing to activate. A node's utility depends on the *size of the connected components of active nodes adjacent to u in G* . Thus, node u activates if the connected component containing u in the subgraph induced in G by nodes $\{v : v \in V, \text{Node } v \text{ is active}\} \cup \{u\}$ has size at least $\theta(u)$. We study the following optimization problem:

Given G and the threshold function $\theta : V \rightarrow \{2, \dots, |V|\}$, what is the smallest feasible seedset $S \subseteq V$ such that if nodes in S activate, then all

remaining nodes in V eventually activate?

This model of node utility captures two key ideas:

1. the traditional notion of “direct network externalities/effects” from economics [40, 68], marketing [13] and other areas [83], that supposes an active node that is part of a network of k active nodes has utility that scales with k , and
2. the fact that we are interested in networking technologies that only allow a pair of active nodes $u, v \in G(V, E)$ to communicate if there is path of active nodes between them in G .

Our model has much in common with the vast literature on diffusion of innovations, and especially the linear threshold model for diffusion in social networks, articulated by Kempe *et al.* [69] and extensively studied in many other works. Indeed, the two models diverge only in the choice of the utility function; ours is non-local, while theirs depends the (weighted) sum of a node’s active *neighbors* in G . Meanwhile, the non-local nature of our utility function has much in common with the classic literature on “direct network externalities/effects” [13, 40, 68, 83] with the important difference that these classic models ignore the underlying graph structure, and instead assume that utility depends on only a *count* of the active nodes. We shall now see that these differences have a substantial effect on our algorithmic results.

5.1.1 Our results.

Our main result is an approximation algorithm based on linear programming that consists of two phases. The first is a linearization phase that exploits combinatorial

properties to encode our problem as an integer program (IP) with a 2-approximate solution, while the second is a randomized rounding algorithm that operates by restricting our search space to *connected seedsets*, *i.e.* seedsets that induce a connected subgraph of G . We have:

Theorem 5.1.1 (Main result). *Consider a networking technology diffusion problem $\{G(V, E), \theta\}$ where the smallest seedset has size opt , the graph has diameter r (*i.e.* r is the length of “longest shortest path” in G), and there are at most ℓ possible threshold values, *i.e.* $\theta : V \rightarrow \{\theta_1, \dots, \theta_\ell\}$. Then there is a polynomial time algorithm that returns a seedset S of size $O(r\ell \log |V| \cdot \text{opt})$.*

Relationship to the linear threshold model in social networks. Our main result highlights the major algorithmic difference between our work and the linear threshold model in social networks [69]. In the social network setting, Chen [20] showed that this problem is devastatingly hard, even when $r, \ell = O(1)$; to avoid this discouraging lower bound, variations of the problem that exploit submodular properties of the objective have been considered (*e.g.* where thresholds are chosen uniformly at random [69] or see [21, 93] and references therein). Indeed, the ubiquity of these techniques seems to suggest that diffusion problems are tractable *only* when the objective exhibits submodularity properties. Our work provides an interesting counterpoint: our positive result does not rely on submodular optimization, and we show that the influence function in our problem, and its natural variations, lacks submodularity properties.

Dependencies on r , ℓ , and $\log |V|$ are necessary. Removing our algorithm’s dependence on r, ℓ , or $\log |V|$ is likely to require a very different set of techniques

because of the following barriers:

1. *Computational barrier.* We use a reduction from Set Cover to show that our problem does not admit any $o(\ln |V|)$ -approximation algorithm, even if $r, \ell = O(1)$.
2. *Information-theoretic barrier.* We present a family of problem instances that prove that any algorithm that returns a connected seedset must pay an $\Omega(r)$ -increase in the size of the seedset in the worst case.
3. *Integrality gap.* The linear program we use has an integrality gap of $\Omega(\ell)$ so that our rounding algorithm is asymptotically optimal in ℓ .

Quality of approximation. We interpret the quality of our approximation for typical problem instances.

Networking. The motivation for our problem is to help centralized authority (*e.g.* a government, a regulatory group) determine the right set of autonomous systems (ASes) in the Internet to target as early adopters for an upgrade to a new networking technology [41, 82]. We comment on the asymptotic order of r and ℓ when a centralized authority executes this algorithm. The graph G is the Internet’s AS-level graph, which is growing over time, with diameter r that does not exceed $O(\log |V|)$ (see, *e.g.* [77]). We remark that the empirical data we have about the Internet’s AS-level topology [8, 23, 36, 99] is the result of a long line of Internet measurement research [94]. On the other hand, obtaining empirical data on ASes’ thresholds is still subject to ongoing research [41, 46]. The following natural assumption and practical constraint restrict the threshold granularity ℓ : (a) ASes should not be sensitive to

small changes in utility (*e.g.* 1000 nodes vs. 1001 nodes), and that (b) in practice, it is infeasible for a centralized authority to obtain information about $\theta(u)$ from every AS u in the Internet, both because this business information is kept private and because, perhaps more importantly, many of these nodes are in distant and possibly uncooperative countries. Thus, thresholds should be chosen from a geometric progression $\{(1 + \epsilon), (1 + \epsilon)^2, \dots, (1 + \epsilon)^\ell\}$ or even restricted to a constant size set $\{5\%, 10\%, 15\%, 20\%, 30\%, 50\%\}$ as in [19, 47, 89] so that $\ell = O(\log |V|)$. Our approximation ratio is therefore polylogarithmic in $|V|$ in this context.

Other settings. Since our model is a general, there could be other settings where ℓ may not be $O(\log |V|)$. Here, the performance of our algorithms is governed by the *stability* of the problem instance. Stability refers to the magnitude of the change in the optimal objective value due to a perturbation of the problem instance, and is commonly quantified using *condition numbers* (as in *e.g.* numerical analysis and optimization [34, 59, 78]). We naturally expect unstable problem instances (*i.e.* with large κ) to be more difficult to solve. Indeed, we can use condition numbers to parameterize our approximation ratio:

Definition 5.1.2 (Condition number). *Consider a problem instance $\Pi = \{G, \theta\}$ and a positive constant ϵ . Let $\Pi^+ = \{G, \theta^+\}$ and $\Pi^- = \{G, \theta^-\}$ be two problem instances on the same graph G where for every $v \in V$, we have $\theta^+(v) = (1 + \epsilon)\theta(v)$ and $\theta^-(v) = (1 - \epsilon)\theta(v)$. Let opt^+ (opt^-) be the value of the optimal solution for Π^+ (Π^-). The condition number is $\kappa(\Pi, \epsilon) \triangleq \frac{\text{opt}^+}{\text{opt}^-}$.*

Corollary 5.1.3. *There exists an efficient algorithm to solve a technology diffusion problem $\Pi = \{G, \theta\}$ whose approximation ratio is $\tilde{O}(\kappa(\Pi, \epsilon) \cdot r)$.*

See details in Chapter 5.4.

Finally, we remark that our IP formulation might also be a promising starting point for the design of new heuristics. Indeed, in Chapter 5.7 we ran a generic IP solver to find seedsets on problem instances of non-trivial size; the seedsets we found were often substantially better than those returned by several natural heuristics (including those used in [10, 19, 47]).

Organization. We present our IP formulation in Chapter 5.2, and describe our rounding algorithm in Chapter 5.3. Lower bounds are in Chapter 5.5. We also present supplementary material on the (lack of) submodularity/supermodularity properties of our problem (Chapter 5.6), our experimental results (Chapter 5.7), and expository examples and figures (Chapter 5.8).

5.2 Linearization & formulating the IP

We now show how to sidestep any potential difficulties that could result from the non-local nature of our setting. To do this, we restrict our problem in a manner that allows for easy encoding using only linear constraints, while still providing a 2-approximation to our objective. We need the following notions:

Activation sequences. Given a seedset S , we can define an *activation sequence* T as a permutation from V to $\{1, \dots, n\}$ where $n = |V|$ that indicates the order in which nodes activate. The t -th position in the sequence is referred as the t -th timestep. We allow a seed node to activate at *any* timestep, while a non-seed node u may activate at a timestep $T(u)$ as long as u is part of a connected component of size at least $\theta(u)$

in the subgraph of G induced by $\{u\} \cup \{v : T(v) < T(u)\}$.

Connected activation sequences. A connected activation sequence T is an activation sequence such that at every timestep t , the set of active nodes induces a connected subgraph of G . We may think of T as a spanning tree over the nodes in the graph, where, at every timestep, we add a new node u to the tree subject to the constraint that u has a neighbor that is already part of the tree.

Our IP will find the smallest seedset S that can induce a connected activation sequence. At first glance this could result in a factor of r growth in the seedset size. However, the following lemma, which may be of independent interest, shows that the seedset size grows at a much smaller rate:

Lemma 5.2.1. *The smallest seedset that can induce a connected activation sequence is at most twice the size of the optimal seedset.*

Notice that requiring the activation sequence T to be connected is *weaker* than requiring a connected seedset S : since T allows a seed to activate *after* a non-seed, the connectivity of T can be preserved by non-seeds whose activation time occurs *between* the activation times of the seed nodes.

Proof of Lemma 5.2.1. Given an optimal activation sequence T_{opt} and seedset opt , we shall transform it into a connected activation sequence T . Along the way, we add nodes to the seedset in manner that increases its size by a factor of at most 2.

Notation. Let $G_i(T)$ be the subgraph induced by the first i active nodes in T . We say a node u is a *connector* in an activation sequence T if the activation of u in T connects two or more disjoint connected components in $G_{T(u)-1}(T)$ into a single component.

Creating a connected activation sequence. Notice that an activation sequence $T(\cdot)$ is connected if and only if there exists no *connector* in the sequence. Thus, it suffices to iteratively “remove” connectors from T until no more connectors remain.

To do this, we initialize our iterative procedure by setting $T \leftarrow T_{\text{opt}}$. Each step of our procedure then finds the earliest connector u to activate in T , adds u to the seedset, and applies the following two transformations (sequentially):

Transformation 1: First, we transform T so that every component in $G_{T(u)}(T)$ is directly connected to u . Let $D(u)$ be the subsequence of T such that every node in $D(u)$ both activates before u , and is part of a component in $G_{T(u)}(T)$ that is *not* connected to u . Transform T so the subsequence $D(u)$ appears immediately *after* node u activates. (This does not harm the feasibility of T , because the nodes in $D(u)$ are disconnected from the other nodes in $G_{T(u)}(T)$ that activate before u .)

Transformation 2: Next, we transform the activation sequence so that it is connected up to time $T(u)$. To see how this works, assume that there are only two connected components C_1 and C_2 in $G_{T(u)-1}(T)$, where $|C_1| \geq |C_2|$. Our transformation is as follows:

1. First, activate the nodes in C_1 as in $T(\cdot)$.
2. Then, activate u . (This does not harm feasibility because we added u to the seedset. Connectivity is ensured because u is directly connected to C_1 .)
3. Finally, have all the nodes in C_2 activate immediately after u ; the ordering of the activations of the nodes in C_2 may be arbitrary as long as it preserves connectivity. (This does not harm feasibility because (a) seed nodes may activate at

any time, and (b) any non-seed $v \in C_2$ must have threshold $\theta(v) \leq |C_2| \leq |C_1|$ and our transformation ensures that at least $|C_1| + 1$ nodes are active before any node in C_2 activates.)

We can easily generalize this transformation to the case where k components are connected by u by letting $|C_1| \geq |C_2| \geq \dots \geq |C_k|$ and activate C_1 , u , and the rest of the components sequentially. At this point, the transformed activation sequence is feasible and connected up to time $t = 1 + |C_1| + |C_2| + \dots + |C_k|$.

Seedset growth. It remains to bound the growth of the seedset due to our iterative procedure. We do this in three steps. First, we observe that the number of extra nodes we added to the seedset is bounded by the number of steps in our iterative procedure. Next, we iteratively apply the following claim (proved later) to argue that the number of steps in our iterative procedure is upper bounded by number of connectors in the optimal activation sequence, T_{opt} :

Claim 5.2.1. *Let T_j be the activation sequence at the start of j^{th} step. The number of connectors in T_{j+1} is less than the number of connectors in T_j .*

Thus, it suffices to bound the number of connectors in T_{opt} . Our third and final step is to show that the number of connectors in T_{opt} is bounded by $|\text{opt}|$. To do this, we introduce a potential function $\Phi(t)$ that counts the number of disjoint connected components in $G_{T_{\text{opt}}(t)}(T)$, and argue the following:

- For every connector u that activates at time t in T_{opt} and joins two or more components, there is a corresponding decrement in Φ , i.e. $\Phi(t) \leq \Phi(t-1) - 1$.
- Next, we have that $\Phi(1) = \Phi(|V|) = 1$, since at the first timestep, there is only

one active node, and at the last timestep all the nodes in the graph are active and form a single giant component. Thus, for every unit decrement in Φ at some time t , there is a corresponding unit increment in Φ at some other time t' .

- Finally, for any unit increment in Φ , *i.e.* $\Phi(t') = \Phi(t' - 1) + 1$, it follows that a new connected component appears in $G_{T_{\text{opt}}(t')}(T)$. This implies that a new seed activates at time t' . Thus, it follows that the number of unit decrements of Φ is upperbounded by the size of the seedset $|\text{opt}|$.

Thus, we may conclude that the number of connectors added to the seedset in our iterative procedure is upperbounded by the number of connectors in T_{opt} which is upperbounded by the size of the optimal seedset opt , and the lemma follows. \square

The correctness of Claim 5.2.1 is fairly intuitive, given that our transformations always preserve the ordering of the nodes that are not in the components joined by node u . We include the proof for completeness.

Proof of Claim 5.2.1. We make use of the following observation:

Observation 1: If two activation sequences T and T' have a common suffix, *i.e.* $T = T'$ for timesteps $\tau, \tau + 1, \dots, |V|$, then T and T' contain the same number of connectors after time $\tau - 1$.

Let $t = T_j(u)$, where u is the earliest connector in T_j . By construction, no connectors exist in T_j prior to time t . Furthermore, we can use Observation 1 to argue that T_j and T_{j+1} contain the same number of connectors after time t . Thus, it

suffices to show that Transformations 1 and 2 in the j^{th} step of our iterative procedure do not introduce new connectors that activate in prior to time t .

Let T^* be the activation sequence after Transformation 1 in the j^{th} step of our iterative procedure, and let $t' = T^*(u)$. We can see that (1) no new connectors activate before time t' in T^* (since, before t' our construction ensures that T^* consists only of active components that are joined by u) and (2) no new connectors activate between time $t' + 1$ and t inclusive (since (a) u was chosen as the earliest connector in T_j , and (b) Transformation 1 preserves the order of the nodes that activate between time $t' + 1$ and t inclusive in T^*).

Finally, we conclude by arguing that Transformation 2 cannot introduce new connectors by (1) applying Observation 1 to the nodes after t' and (2) observing that after Transformation 2, the nodes that activate before t' create a single connected component, and thus by definition cannot contain any connectors. \square

IP encoding. The beauty of a connected activation sequence T is that every nonseed node's decision to activate becomes local, rather than global: node v need only check if (a) at least one of its neighbors are active, and (b) the current timestep t satisfies $t \geq \theta(v)$. Moreover, given a connected activation sequence T , we can uniquely recover the smallest feasible seedset S that could induce T by deciding that node u is a seed iff $\theta(u) > T(u)$. Thus, our IP encodes a connected activation sequence T , as a proxy for the seedset S . Let $x_{i,t}$ be an indicator variable such that $x_{i,t} = 1$ if and only if $T(v_i) = t$. The integer program is presented in Figure 5.1. The *permutation constraints* guarantee that the variables $x_{i,t}$ represent a permutation. The *connectivity constraints* ensure that if $x_{i,t} = 1$ (i.e. node v_i activates at step t), there is some other

node $u_{i'}$ such that $v_{i'}$ (a) is a neighbor of node v_i and (b) activates at earlier time $t' < t$. Finally, the objective function minimizes the size of the seedset by counting the number of $x_{i,t} = 1$ such that $t < \theta(v_i)$.

We remark that our IP formulation suggests a similarity between our setting and the vehicle routing with time windows problem (*e.g.* [11, 15, 35, 44]). Consider a time windows problem, where we are given an undirected metric graph G and time window $[r(u), d(u)]$ for each node u , and our objective is to choose a tour for the vehicle through G that visits as many nodes as possible during their respective time windows. In our setting (restricted to connected activation sequences), the tour becomes a spanning tree, and each node u has time window $[\theta(u), n]$. Understanding the deeper connection here is an interesting open question.

5.3 Rounding algorithm.

Unfortunately, the simple IP of Figure 5.1 has a devastating $\Omega(n)$ integrality gap (Chapter 5.5.3 presents a negative example). We eliminate this integrality gap by

	min	$\sum_{i \leq n} \sum_{t < \theta_i} x_{i,t}$	
subject to:	$\forall t, i :$	$x_{i,t} \in \{0, 1\}$	
	$\forall i :$	$\sum_{t \leq n} x_{i,t} = 1$	(permutation constraints)
	$\forall t :$	$\sum_{i \leq n} x_{i,t} = 1$	(permutation constraints)
	$\forall t > 1, i :$	$\sum_{\{v_i, v_{i'} \in E\}} \sum_{t' < t} x_{i',t'} \geq x_{i,t}$	(connectivity constraints)

Figure 5.1: Simple IP for the networking technology diffusion problem.

adding extra constraints to the IP of Figure 5.1, and refer to the resulting IP as the *augmented IP*. We defer presentation of this IP to Chapter 5.3.2 and focus now on the high level structure of our rounding algorithm.

Our rounding algorithm is designed to exploit the relationship between seedset S and connected activation sequences T ; namely, the fact that we can uniquely recover a S from T by deciding that node u is a seed if $T(u) < \theta(u)$. As such, it returns *both* S and T with the following four properties:

1. *Consistency.* S and T are *consistent*; namely, T is an activation sequence for the diffusion process induced by $\{G, \theta, S\}$. (Recall that T is such that any seed $u \in S$ can activate at any time, and any non-seed $u \notin S$ can activate whenever it is connected to an active component of size at least $\theta(u) - 1$.)
2. *Feasibility.* T is such that every node eventually activates.
3. *Connectivity.* T is a connected activation sequence.
4. *Small seedset.* The seedset S has small size, *i.e.* size bounded in the objective function of our LP.

But how should we round the fractional $x_{i,t}$ values returned our LP relaxation to achieve this? Let's first consider two natural approaches for sampling S and T :

Approach 1: Sample the seedset S : Recall that in a connected activation sequence, a node that activates at time $t < \theta(u)$ must be a seed. Therefore, we can sample the seedset S by adding each node u_i to S with probability proportional to

$$\sum_{t < \theta(u_i)} x_{i,t}.$$

Approach 2: Sample the activation sequence T : We can instead sample the activation sequence T by deciding that node u_i activates before time t with probability proportional to $\sum_{\tau < t} x_{i,\tau}$.

However, neither of these approaches will work very well. While Approach 1 guarantees that the seedset S is small (Property 4), it completely ignores the more fine-grained information provided by the $x_{i,t}$ for $t \geq \theta(v_i)$ and so its not clear that nonseed nodes will activate at the right time (Property 2). Meanwhile, Approach 2 guarantees feasibility (Property 2), but by sampling activation times for each node independently, it ignores correlations between the $x_{i,t}$. It is therefore unlikely that the resulting T is connected (Property 3), and we can no longer extract a small seedset (Property 4) by checking if $T(u) < \theta(u)$.

Instead, we design a sampling procedure that gives us a coupled pair $\{S, T\}$ where, with high probability, (a) the distribution of S will be similar to that of Approach 1, so that the seedset is small (Property 4), while (b) the distribution of T will be similar to Approach 2, so we have feasibility (Property 2), and also (c) that T is connected (Property 3). However, S and T are not necessarily consistent (Property 1). Later, we show how we use repeated applications of the sampling approach below to “error-correct” inconsistencies, but for now, we start by presenting the sampling routine:

Approach 3: Coupled sampling. We start as in Approach 1, adding each $v \in V$ to S with probability $\min \left\{ 1, \alpha \sum_{t < \theta(v)} x_{v,t} \right\}$. We next run *deterministic* processes: $S \leftarrow \text{GLUE}(S)$ followed by $T \leftarrow \text{GETSEQ}(S)$.

The GLUE procedure, defined below, ensures that S is connected (*i.e.* induces

a connected subgraph of G), and blows up $|S|$ by an $O(r)$ -factor (where r is graph diameter). Meanwhile, GET-SEQ, defined in Chapter 5.3.2, returns a connected activation sequence T ; we remark that T may not be a permutation (many nodes or none could activate in a single timestep), and may not be feasible, *i.e.* activate every node.

GLUE(S)

```

1  while  $S$  is not connected
2      do
3          Let  $C$  be a connected component in the subgraph induced by  $S$ .
4          Pick  $u \in S \setminus C$ . Let  $P$  be the shortest path connecting  $u$  and  $C$  in  $G$ .
5          Add nodes in  $P$  to  $S$ .
6  return  $S$ .
```

The properties of Approach 3 are captured formally by the following proposition, whose proof (Chapter 5.3.2) presents a major technical contribution of our work:

Proposition 5.3.1. *Let $\alpha = 24(1+\epsilon) \ln(\frac{4n^2}{\epsilon})$ and ϵ be a suitable constant. Then there exists an augmented IP and an efficiently computable function GET-SEQ(\cdot) such that Approach 3 returns S and T (that are not necessarily consistent), where*

1. T is connected,
2. for any $v \notin S$ we have that $T(v) \geq \theta(v)$, and
3. for any v and t ,

- if $\sum_{t' \leq t} x_{v,t'} \geq \frac{1}{12(1+\epsilon)}$, then $\Pr[T(v) \leq t] \geq 1 - \frac{\epsilon}{4n^2}$.

- if $\sum_{t' \leq t} x_{v,t'} < \frac{1}{12(1+\epsilon)}$, then $\Pr[T(v) \leq t] \geq (1 + \epsilon)(\sum_{t' \leq t} x_{v,t'})$.

Notice that the third item in Proposition 5.3.1 suggests that the distribution of T in Approach 3 is “close” to that of Approach 2. In what follows, we apply the ideas we developed thus far to design an algorithm that uses Proposition 5.3.1 to “error-correct” inconsistencies between S and T so that all four properties are satisfied. Then, in Chapter 5.3.2 we present the more technically-involved proof of Proposition 5.3.1.

5.3.1 Resolving inconsistencies using rejection-sampling

Recall the threshold function $\theta : V \rightarrow \{\theta_1, \dots, \theta_\ell\}$, and suppose a threshold θ_j is *good* with respect to T if there are at least $\theta_j - 1$ active nodes in T by time $\theta_j - 1$. The following simple lemma presents the properties we need from our rejection sampling algorithm:

Lemma 5.3.2. *Let S be a seedset and T be an activation sequence. If*

- (P1). *T is connected and feasible (for any $v \in V$, $T(v) \leq n$), and*
- (P2). *$T(v) \geq \theta(v)$ for all $v \notin S$, and*
- (P3). *every θ_j for $j \in [\ell]$ is good with respect to $T(\cdot)$,*

then S is consistent with T and S is a feasible seedset.

Proof. To show that S and T are consistent, we argue that by the time a non-seed $v \notin S$ activates in T , there are at least $\theta(v) - 1$ active nodes. Since v activates at time $T(v) \geq \theta(v)$, this follows because $T(\cdot)$ is connected and each θ_j is good. Since T is feasible and S is consistent with T , we have that S is feasible. \square

We construct a pair of $\{S, T\}$ that meets the properties of Lemma 5.3.2 in two phases. First, we construct ℓ pairs $\{S_1, T_1\}, \dots, \{S_\ell, T_\ell\}$ where for each $\{S_j, T_j\}$ we have that (P.1) and (P.2) hold, a *single* threshold θ_i is good w.r.t. T_i , and S_i is “small”, *i.e.* $|S_i| \leq 24(1 + \epsilon)^2 \ln(\frac{4n^2}{\epsilon})r \cdot \text{opt}$. The second phase assembles these ℓ pairs into a single $\{S, T\}$ pair so *all* θ_j are good w.r.t. T , so that (P1)-(P3) hold, and the seedset S is bounded by $O(r\ell \ln n \cdot \text{opt})$, so our main result follows.

Step 1. Rejection-sampling to find $\{S_j, T_j\}$ pairs $\forall j \in [\ell]$. The following lemma shows that we can repeat Approach 3 until we find S_j, T_j that satisfy the properties above:

Lemma 5.3.3 (Success of a single trial). *Let S_j and T_j be sampled as in Approach 3. For any t , let A_t be the number of nodes active in T_j up to time t (inclusive). Then $\Pr[A_t \geq t \wedge A_n = n] \geq \frac{\epsilon}{2n}$.*

To see why, observe that (P1)-(P2) hold by Proposition 5.3.1, and θ_j is good w.r.t. T_j with probability $\frac{\epsilon}{2n}$ by Lemma 5.3.3, and S_j has the required size with probability $\geq 1 - \frac{1}{n^{10}}$ by standard Chernoff bounds (the exponent 10 here is chosen arbitrarily). Therefore, we successfully find the required $\{S_j, T_j\}$ with probability $\frac{\epsilon}{2n} - \frac{1}{n^{10}}$ in a single trial. After $O(n \log n)$ independent trials, we find the required $\{S_j, T_j\}$ with probability $1 - 1/n^c$ for sufficiently large c .

Proof of Lemma 5.3.3. Recall that A_t is number of active nodes by time t (inclusive). We have

$$\Pr[A_n < n] = \Pr[\exists v : T(v) > n] \leq \sum_{v \in V} \Pr[T(v) > n] \leq \frac{n\epsilon}{4n^2} = \frac{\epsilon}{4n}. \quad (5.1)$$

The last inequality holds because of Proposition 5.3.1. It suffices to show that $\Pr[A_t \geq t] \geq \frac{3\epsilon}{4n}$ since $\Pr[A_n = n \wedge A_t \geq t] \geq \Pr[A_t \geq t] - \Pr[A_n < n]$.

Let us partition V into heavy nodes H , and light nodes L . We put $v \in H$ when $\sum_{\tau \leq t} x_{v,\tau} \geq \frac{1}{12(1+\epsilon)}$, and $v \in L$ otherwise. Let's consider two cases, based on the “weight” of the light nodes ρ_t :

$$\rho_t = \sum_{v \in L} \sum_{\tau \leq t} x_{v,\tau} \quad (5.2)$$

Case 1. $\rho_t < 1$ (*The light nodes are very light*). Recalling that the permutation constraints of our LP impose that $\sum_{v \in V} \sum_{\tau < t} x_{v,\tau} = t$, it follows that

$$t - 1 < t - \rho_t = \sum_{v \in V} \sum_{\tau < t} x_{v,\tau} - \sum_{v \in L} \sum_{\tau \leq t} x_{v,\tau} = \sum_{v \in H} \sum_{\tau \leq t} x_{v,\tau} \leq t$$

Using the first and last inequalities and taking the ceiling, we get that

$$|H| \geq \left\lceil \sum_{v \in H} \sum_{\tau \leq t} x_{v,\tau} \right\rceil = t.$$

Since $|H| \geq t$, if every node in H activates before time t we know that $A_t \geq t$. We write

$$\Pr[A_t \geq t] \geq \Pr[T(v) \leq t, \forall v \in H] \geq 1 - \sum_{v \in H} \Pr[T(v) > t] \geq 1 - \frac{\epsilon}{4n}, \quad (5.3)$$

where the last inequality in (5.3) holds because of Proposition 5.3.1.

Case 2. $\rho_t \geq 1$ (*The light nodes are not very light*). We start by defining two events.

\mathcal{E}_1 is the event that all the heavy nodes are active by time t , i.e. $T(v) \leq t \forall v \in H$.

\mathcal{E}_2 is the event that at least ρ_t light nodes are on by time t , i.e. $|\{v \in L \wedge T(v) \leq t\}| > \rho_t$.

When both \mathcal{E}_1 and \mathcal{E}_2 occur, we have

$$A_t \geq |H| + \rho_t \geq \sum_{v \in H} \sum_{\tau \leq t} x_{v,\tau} + \sum_{v \in L} \sum_{\tau \leq t} x_{v,\tau} = t$$

where both the second inequality and the last equality use the permutation constraints of the LP. It follows that $\Pr[A_t > t] \geq \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2] \geq \Pr[\mathcal{E}_2] - \Pr[\neg \mathcal{E}_1]$. We now bound each event individually.

Let's start by bounding $\Pr[\mathcal{E}_2]$. Letting $I(\cdot)$ be an indicator variable that sets to 1 iff the parameter is true, we have that

$$\mathbb{E}[\sum_{v \in L} I(T(v) \leq t)] = \sum_{v \in L} \Pr[T(v) \leq t] \geq \sum_{v \in L} \left((1 + \epsilon) \sum_{t' \leq t} x_{v,t'} \right) = (1 + \epsilon) \rho_t \quad (5.4)$$

where the inequality uses Proposition 5.3.1 as usual. Meanwhile, using the law of total probability we get

$$\mathbb{E}[\sum_{v \in L} I(T(v) \leq t)] \leq \Pr[\mathcal{E}_2]n + \Pr[\neg \mathcal{E}_2]\rho_t \leq \Pr[\mathcal{E}_2]n + \rho_t \quad (5.5)$$

Combining (5.4)-(5.5) we find that $\Pr[\mathcal{E}_2] \geq \frac{\epsilon \rho_t}{n} \geq \frac{\epsilon}{n}$. Next, we bound $\Pr[\mathcal{E}_1]$ by observing that

$$\Pr[\neg \mathcal{E}_1] \leq \sum_{v \in H} \Pr[T(v) > t] \leq \frac{\epsilon}{4n}$$

using Proposition 5.3.1 for the last inequality again. Finally, we combine both bounds to conclude that $\Pr[A_t > t] \geq \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2] \geq \Pr[\mathcal{E}_2] - \Pr[\neg \mathcal{E}_1] \geq \frac{3\epsilon}{4n}$ as required. \square

Step 2. Combine the $\{S_i, T_i\}$ to obtain the final $\{S, T\}$. We can now construct our final $\{S, T\}$ pair in a rather straightforward way: to construct S , we take the union of all the S_j 's and then use GLUE to connect them; that is we take $S \leftarrow \text{GLUE}(\bigcup_{j \leq \ell} S_j)$. To construct T , we set $T(v) = 1$ for all seeds $v \in S$ and $T(v) = \min_{j \leq \ell} T_j(v)$ ($\forall v \in V \setminus S$).

To conclude, we need show that this $\{S, T\}$ pair satisfies Lemma 5.3.2.

- First we show (P1) holds. Since every T_j is feasible, and $T(v) \leq T_j(v)$ by construction, it follows that T is also feasible. Next we show that $T(v)$ is connected by induction over t . As a base case, observe that $T = \min_{j \leq \ell} T_j(v)$ is connected at $t = 1$, since the seedset $S = \text{GLUE}(\bigcup_j S_j)$ is connected. As the induction step, we assume that T is connected up to time t (inclusive) and show that T is also connected up to time $t + 1$ (inclusive). To do this, let v be a node such that $T(v) = t + 1$. It follows that there exists $j \leq \ell$ such that $T_j(v) = t + 1$; since T_j is connected, there must be another node u such that $T_j(u) < t + 1$ and u and v are neighbors in the graph G . Since $T(u) \leq T_j(u)$, it follows that v is connected to a node (namely node u) that is active at time $t + 1$, and the induction step follows.
- We show that (P2) holds. For all $v \notin S$, we have $v \notin S_j$ for all $j \leq \ell$. This means $T_j(v) \geq \theta(v)$ for all j . Therefore, $T(v) \geq \theta(v)$ and (P2) holds.
- Finally, (P3) holds. For each $j \leq \ell$ we know that θ_j is good w.r.t to T_j . For all $j \leq \ell$, every node v has $T(v) \leq T_j(v)$ by construction, so that the number of active nodes at time θ_j in T must be no fewer than the number of active nodes in T_j . (P3) follows since θ_j is good w.r.t to T_j for every $j \leq \ell$.

It follows that Lemma 5.3.2 holds and the final seedset S is indeed a feasible seedset. Since the size of each seedset S_i is bounded by $O(r \log n \cdot \text{opt})$ (and the gluing in Phase 2 grows the seedset by an additive factor of at most $\ell \cdot r$) it follows that S has size at most $O(\ell r \log n \cdot \text{opt})$ and our main result follows.

$t = 1$	$x_{A,1} = 0.1$
$t = 2$	$x_{B,2} = 0.1$ (because $x_{B,2} \leq x_{A,1}$)
$t = 3$	$x_{C,3} = 0.1$ (because $x_{C,3} \leq x_{B,1}$)
$t = 4$	$x_{B,4} = 0.2$ (because $x_{B,4} \leq x_{A,1} + x_{C,3}$)
$t = 5$	$x_{C,5} = 0.2$ (because $x_{C,5} \leq x_{B,2} + x_{B,4}$)
$t = 6$	$x_{B,6} = 0.4$ (because $x_{B,6} \leq x_{A,1} + x_{C,3} + x_{C,5}$)
$t = 7$	$x_{C,7} = 0.7$ (because $x_{C,7} \leq x_{B,2} + x_{B,4} + x_{B,6}$)
$t = 8$	$x_{B,8} = 0.3$ (because $x_{B,8} \leq x_{A,1} + x_{C,3} + x_{C,5} + x_{C,7}$)
$t = 9$	$x_{A,9} = 0.9$ (because $x_{A,9} \leq \sum_{t' \leq 8} x_{B,t'}$)

Figure 5.2: A pathological example for the simple LP.

5.3.2 Strengthened IP and coupled sampling

We show how we use a flow interpretation of our problem to prove Proposition 5.3.1.

The need for stronger constraints

In Chapter 5.5.3, we show that the LP in Figure 5.1 has an $\Omega(n)$ integrality gap. To understand why this gap comes about, let us suppose that each $x_{i,t}$ returned by the LP is a mass that gives a measure of the probability that node v_i activates at time t . Consider the following example (also see Figure 5.2):

Pathological example. Consider a graph that contains a clique of nodes A,B and C. Suppose the LP returns a solution such that at $t = 1$, node A has mass 0.1, while all other nodes have mass 0. The constraints repeatedly allow mass from node A to circulate through nodes B and C and then back to A, as shown in the variable assignments beside. Finally, at $t = 9$, enough mass has circulated back to A, so that A has mass 0.9 and thus “probability” 0.9 of activating. Note that this is highly artificial, as all of this mass originated at A to begin with! In fact, no matter how we

interpret these $x_{i,t}$, the example suggests that this “recirculation of mass” is unlikely to give us any useful information about when node A should actually activate.

The flow constraints

As the example above suggests, we can think of the diffusion process in the context of network flows. Specifically, we suppose that when a nonseed node u activates at time $T(u)$, a unit flow originates at a seed node and flows to node u along the network induced by the nodes active prior to timestep $T(u)$. We therefore augment the IP of Figure 5.1 with this idea by introducing *flow constraints*:

The flow network. For any solution $\{x_{i,t}\}_{i,t \leq n}$, we define a *flow network* \mathcal{H} , in with vertex set $V(\mathcal{H}) = \{X_{i,t} : i, t \in [n]\}$ and edge set $E(\mathcal{H}) = \{(X_{i,t}, X_{i',t'}) : t' > t \wedge \{v_{i'}, v_i\} \in E(G)\}$. Every node $X_{i,t}$ in the flow network \mathcal{H} has capacity $x_{i,t}$, while edges in \mathcal{H} do not have capacity bounds. We let the line that connects nodes $\{X_{i,t} : t = \theta(v_i)\}$ be the *threshold line*. All the $X_{i,t}$ such that $t < \theta(v_i)$ are flow graph nodes *to the left of the threshold line*; very roughly, these nodes corresponds the region where v_i is a seed. The rest are flow graph nodes *to the right of the threshold line*, and roughly correspond to v_i being a nonseed. A sample flow graph and its threshold line appears in Chapter 5.8.

Flow constraints. For now, we suppose the first node to be activated in the optimal solution is known to be v_1 (so that $x_{1,1} = 1$); we can see later this assumption can be removed. For any i and $t \geq \theta(v_i)$, we define the (i, t) -flow as the multiple-sink flow problem over the flow network \mathcal{H} , where the source is $X_{1,1}$ and the sinks are nodes to the right of threshold line, namely $\{X_{i,\theta(v_i)}, X_{i,\theta(v_i)+1}, \dots, X_{i,t}\}$. The demand

for the sink $X_{i,t}$ is $x_{i,t}$. Our flow constraints require that every (i, t) -flow problem (for all i and all $t \geq \theta(v_i)$) has a solution.

We have the following Lemma.

Lemma 5.3.4. *The augmented IP for the technology diffusion problem is such that*

- *when $T(v_1) = 1$ in the optimal connected activation sequence, this IP returns the same set of feasible solutions as the simple IP of Figure 5.1.*
- *the fractional solution for the corresponding relaxed LP satisfies all the (i, t) -flow constraints.*

Furthermore, we can solve the corresponding relaxed linear program efficiently.

Proof of Lemma 5.3.4. Our implementation is presented in Figure 5.3. We prove Lemma 5.3.4 in three parts. First, we show that if we add following two constraints to the IP in Figure 5.1: (a) $x_{1,1} = 1$ and (b) that (i, t) -flow problems have feasible solutions for all i and $t \geq \theta(v_i)$, then the resulting IP returns the subset of solutions of the original IP where $T(v_1) = 1$. We also remark on how to remove the assumption that $T(v_1) = 1$ in the optimal T . Second, we show how to encode the flow constraints as an IP. Finally, we mention why the corresponding relaxed LP is efficiently solvable.

Part 1. Any connected activation sequence satisfies the flow constraints It suffices to show that for any connected activation sequence, its corresponding integral variables $\{x_{i,t}\}_{i,t \leq n}$ satisfy the (i, t) -flow constraints for all i and $t \geq \theta(v_i)$.

In what follows, we both use $\{x_{i,t}\}_{i,t \leq n}$ and $T(\cdot)$ to represent the activation sequence. Let $\{x_{i,t}\}_{i,t \leq n}$ be a connected activation sequence. Let us consider an

min	$\sum_{i \leq n} \sum_{t < \theta_i} x_{i,t}$	
subject to:		
$\forall i, t :$	$x_{i,t} \in \{0, 1\}$	
$\forall i$	$\sum_{t \leq n} x_{i,t} = 1$	(permut'n cstr)
$\forall t$	$\sum_{i \leq n} x_{i,t} = 1$	(permut'n cstr)
$\forall t > 1, i :$	$\sum_{\{v_i, v_{i'} \in E\}} \sum_{t' < t} x_{i',t'} \geq x_{i,t}$	(connectivity cstr)
	$x_{1,1} = 1$	(make $X_{1,1}$ the source)
$\forall i, t \geq \theta(u_i)$		
\forall partitions of $V(\mathcal{H})$		
S, \bar{S} , s.t. $X_{1,1}^+ \in S, \text{sk}_{i,t} \in \bar{S} \quad \sum_{e \in \delta(S, \bar{S})} c(e) \geq \sum_{\theta(u_i) \leq t' \leq t} x_{i,t'}$ (flow cstr).		

Figure 5.3: Integer program for solving the technology diffusion problem.

arbitrary (i, t) -flow. Let τ be the time step such that $x_{i,\tau} = 1$. Recall that the demand in an (i, t) -flow problem is $\sum_{\theta(u_i) \leq \tau \leq t} x_{i,\tau}$. Therefore, when $\tau > t$ or $\tau < \theta(v_i)$, the demand is 0 and we are done. We only need to consider the case where $\theta(v_i) \leq \tau \leq t$. We claim that when $\{x_{i,t}\}_{i,t \leq n}$ is a connected activation sequence, for any t and $v_k \triangleq T^{-1}(t)$, there exists a path $v_1 v_{i_1} v_{i_2} \dots v_{i_j} v_k$ such that

$$T(v_1) < T(v_{i_1}) < \dots < T(v_{i_{j-1}}) < T(v_{i_j}) < T(v_k) = t.$$

This can be seen by induction on t . For the base case, $t = 2$ and the path is $v_1 v_k$. For the induction step, suppose the claim holds for every time step up to $t - 1$. We show that it also holds when v_k activates at the t -st time step. Since $\{x_{i,t}\}_{i,t \leq n}$ is connected, there exists a $v_{k'}$ such that there is an edge $\{v_{k'}, v_k\} \in E$ and $T(v_{k'}) < T(v_k)$. By the induction hypothesis, there must be a path $v_1 \dots v_{k'}$ that connects v_1 and $v_{k'}$, where the activation time of each node on the path increases monotonically. Thus, the path we seek is $v_1 v_{i_1} \dots v_{k'} v_k$, which completes the proof of the induction step.

We conclude the proof by using the claim we proved by induction. Namely, there is a path from v_1 to the node v_k activating at time t . It follows that we can push a unit of flow along the path induced in the flow graph \mathcal{H} , namely $X_{1,1}, X_{i_1, T(v_{i_1})}, \dots, X_{i_j, T(v_{i_j})}, X_{k,t}$, so we must have a feasible solution to the (i, t) -flow problem.

Turning on v_1 . We remark that while we have been assuming that v_1 is known to activate at $t = 1$ in the optimal solution, we can ensure this assumption holds by polynomial-time “guessing”; run the IP $O(|V|)$ times, relabeling a different node in the graph as v_1 in each run, and use the run that returns the smallest seedset.

Part 2. Implementation of the flow constraints. The (i, t) -flow constraints are enforced via the max-flow-min-cut theorem, *i.e.* by using the fact that the minimum

cut between the source and the sinks is the same as the maximum flow. Thus, to ensure every (i, t) -flow problem has a feasible solution, we require the capacity for all the cuts between the source and the sinks to be larger than the demand. The actual implementation is quite straightforward, but we present the details of the IP for completeness:

- The capacity constraints we have are over the nodes in \mathcal{H} . We use standard techniques to deal with this: we replace each node $X_{i,t}$ in \mathcal{H} with two nodes $X_{i,t}^+$ and $X_{i,t}^-$ connected by a directed edge of capacity $x_{i,t}$.
- There are multiple sinks in a (i, t) -flow problem. To deal with this, for every i and $t \geq \theta(v_i)$, we introduce a new node $\text{sk}_{i,t}$ to \mathcal{H} that is connected to every sink $X_{i,\theta(v_i)}, X_{i,\theta(v_i)+1}, \dots, X_{i,t}$ that sinks *all* the flow in the (i, t) -flow problem.

Our implementation is presented in Figure 5.3. Let S and \bar{S} be two arbitrary partition of the nodes in \mathcal{H} . We let $\delta(S, \bar{S})$ be the cut of the partition, *i.e.* the set of edges whose end points are in different subsets of the partition. Also, we let $c(e)$ be the capacity of the edge e , *i.e.* $c(\{X_{i,t}^+, X_{i,t}^-\}) = x_{i,t}$ and $c(e) = \infty$ for all other edges.

Part 3. The relaxed linear program is efficiently solvable. Our relaxed LP contains an exponential number of constraints (namely, the flow constraints). Nevertheless, we can use the ellipsoid method to find an optimal solution in polynomial time using a separation oracle [110] that validates if each of the (i, t) -flow problems over \mathcal{H} have solutions, and if not, returns a min-cut constraint that is violated. This oracle can be constructed using algorithms in, *e.g.* [54]. □

Eliminating the integrality gap. The flow constraints eliminate the pathological

example above, and therefore also the $\Omega(n)$ integrality gap. To see why, notice that the $(B, 4)$ -flow problem has total demand 0.2 (*i.e.* $x_{B,2} = 0.1$ and $x_{B,4} = 0.1$) but there is no way to supply this demand from $X_{A,1}$.

Why coupled sampling works.

In addition to improving the robustness of our IP, the flow constraints also have the following pleasant interpretation that we use in the design our rounding algorithm: if there is a flow $f \in [0, 1]$ from a seed node to a non-seed node u at time t , then node u has probability f of activating at time t .

On connected seedsets. To ensure that all network flows originate at seed nodes, Approach 3 requires GET-SEQ to return an activation sequence T where all seed nodes activate before the non-seed nodes. If we couple this with the requirement that T is connected (so we can use the trick of deciding that node v is a seed if $T(v) < \theta(v)$), it follows that we require a *connected seedset* S (*i.e.* the nodes in S induce a connected subgraph of G). Approach 3 achieves this by using GLUE to connect the nodes it samples into its seedset S , and then deterministically generates T using GET-SEQ as specified below.

GET-SEQ(\mathcal{H}, S)

- 1 Initialize by flagging each $X_{u,t} \in \mathcal{H}$ as “inactive” by setting $b_{u,t} \leftarrow 0$.
- 2 $\forall u \in S, b_{u,t} \leftarrow 1$ for $t < \theta(u)$.
- 3 // “Activate” each $X_{u,t}$ to the left of the threshold line
- 4 **for** $t \leftarrow 1$ to n
- 5 **do** $\forall u$ s.t. $\theta(u) \geq t$:
- 6 **if** $(\exists v, \tau : ((X_{v,\tau}, X_{i,t}) \in E(\mathcal{H})) \wedge (b_{v,\tau} = 1))$
- 7 $b_{u,t} \leftarrow 1$ for $t \geq \theta(u)$
- 8 // “Activate” each $X_{u,t}$ to the right of the threshold line
- 9 Obtain T by taking $T(u) \leftarrow \min \{t : b_{u,t} = 1\}$ for every $u \in V$.
- 10 **return** T .

Intuition behind the proof of Proposition 5.3.1. Given the probabilistic interpretation of flows, consider what happens if two disjoint flows f_1 and f_2 originate from different seeds and arrive simultaneously at node u at time t . The total flow at node u at time t is then $f_1 + f_2$. What does this merge of two disjoint flows mean in our probabilistic interpretation? It turns out that the natural interpretation is already pretty sensible: with probability f_1 , the technology is diffused via the first flow, and with probability f_2 the technology is diffused via the second flow. Now, the probability that the technology is diffused to u via either of these two flows is $1 - (1 - f_1)(1 - f_2)$. When f_1, f_2 are both small, this probability becomes $\approx f_1 + f_2$, so that the total flow can be used to determine node u ’s activation probability. On the other hand, when f_1 or f_2 is large, we are fairly confident that u should activate prior to time t , and so we can simply decide that $T(u) \leq t$ without incurring a large

increase in the size of the seedset. Given that the total demand in the (u, t) flow problem is $\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$, it follows that the probability that u is a nonseed and is activated by time t is roughly proportional to this demand. Also, notice that u itself is chosen as a seed with probability $\sum_{\tau < \theta(u)} x_{u,\tau}$ so by combining these events in the appropriate way, we get that $\Pr[T(u) \leq t] \propto \sum_{\tau < t} x_{u,\tau}$ as required by the third item in Proposition 5.3.1.

To formalize this intuition, we need a few definitions. First, for each pair u and t (where $t \geq \theta(u)$), let an arbitrary (but fixed) solution $\mathcal{F}_{u,t}$ for the (u, t) -flow problem be the *representative flow* for the (u, t) -flow problem. To help us understand how disjoint flows merge, we use the following notion:

Definition 5.3.5 (Border nodes). *Consider the (u, t) -flow problem on the flow graph \mathcal{H} and the corresponding representative flow $\mathcal{F}_{u,t}$. Let us decompose the flow into paths (in an arbitrary but consistent manner) $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_q$. Consider an arbitrary path \mathcal{P}_k and let $X_{j,\tau}$ be the last node on \mathcal{P}_k that is to the left of the threshold line. Define $X_{j,\tau} = \text{border}(\mathcal{P}_k)$.*

- *The border nodes for the (u, t) -flow problem on flow graph \mathcal{H} are $\beta(u, t) \triangleq \{\text{border}(\mathcal{P}_1), \dots, \text{border}(\mathcal{P}_q)\}$.*
- *The border nodes for the (u, t) -flow problem on G are*

$$B(u, t) \triangleq \{v_j : \exists \tau \text{ s.t. } X_{j,\tau} \in \beta(u, t)\}.$$

For notational convenience, when $t < \theta(u)$, we let $\beta(u, t) = B(u, t) = \emptyset$.

An expository example of G , \mathcal{H} and their border nodes is in Chapter 5.8. Border nodes are useful because GET-SEQ ensures any nonseed node u activates before time

$t > \theta(u)$ whenever a border node in $B(u, t)$ is in the seedset S . Letting p_j be the probability that node v_j is placed in the S in a single run of Approach 3, and defining the *seed weight* of node v_j as $\omega_j \triangleq \sum_{t < \theta(v_j)} x_{j,t}$ so that $p_j = \min\{1, \alpha\omega_j\}$ (recall that α is our sampling bias in Approach 3), it follows that $\Pr[T(u) \leq t]$ is related to $\sum_{v_j \in B(u,t)} \omega_j$. The following lemma therefore allows us to relate $\Pr[T(u) \leq t]$ to the demand in the (u, t) -flow problem $\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$, which is the main task of the proof of Proposition 5.3.1:

Lemma 5.3.6 (Border node lemma). $\sum_{v_j \in B(u,t)} \omega_j \geq \sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$ for any $u \in V$ and $t \geq \theta(u)$.

This lemma uses the fact the demand of the (u, t) -flow problem is upperbounded by the total capacity of the border nodes $B(u, t)$, which is in turn upperbounded by the total seed weight of the border nodes.

Proof of Lemma 5.3.6. Let us decompose the representative flow $\mathcal{F}_{i,t}$ into paths (in an arbitrary but consistent manner) $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_q$, and let f_k be the volume of the flow on path P_k .

$$\begin{aligned}
\sum_{\theta(v_i) \leq \tau \leq t} x_{i,\tau} &= \sum_k f_k && \text{(the demand in the } (i, t) \text{ flow problem is satisfied)} \\
&= \sum_{X_{j,\tau} \in \beta(i,t)} \sum_{\text{border}(P_k)=X_{j,\tau}} f_k && \text{(multiple border}(P_k)\text{ can map to a single } X_{j,\tau}) \\
&\leq \sum_{X_{j,\tau} \in \beta(i,t)} x_{j,\tau} && \text{(bounding capacity of } X_{j,\tau}) \\
&= \sum_{v_j \in B(i,t)} \sum_{\tau \text{ s.t. } X_{j,\tau} \in \beta(i,j)} x_{j,\tau} && \text{(translating from } \mathcal{H} \text{ to } G) \\
&\leq \sum_{v_j \in B(i,t)} \sum_{\tau \leq \theta(v_j)} x_{j,\tau} && (\tau \text{ s.t. } X_{j,\tau} \in \beta(i,j) \Rightarrow \tau \leq \theta(v_j)) \\
&= \sum_{v_j \in B(i,t)} w_j && \text{(definition of } w_j)
\end{aligned}$$

Notice that the last four lines give the total seed weight of the border nodes as an upper bound on their total capacity. \square

Armed with our border node lemma, we can move on to our main task:

Proof of Proposition 5.3.1. One can verify, by the construction of GETSEQ, that the activation function T is always connected and for any $u \notin S$, $T(u) \geq \theta(u)$. Our main objective here is to prove that $\Pr[T(u) \leq t] \propto \sum_{\tau < t} x_{i,t}$ for every pair (u, t) where $u \in V$ and $t \leq n$. More specifically, we need to show that:

Part 1. If $\sum_{t' \leq t} x_{u,t'} < \frac{1}{12(1+\epsilon)}$, then $\Pr[T(u) \leq t] \geq (1 + \epsilon)(\sum_{t' \leq t} x_{u,t'})$ (5.6)

Part 2. If $\sum_{t' \leq t} x_{u,t'} \geq \frac{1}{12(1+\epsilon)}$, then $\Pr[T(u) \leq t] \geq 1 - \frac{\epsilon}{4n^2}$.

Our proof relies on the observation that $T(u) \leq t$ if at least one of the following events hold:

\mathcal{E}_1 : u is seed (because GETSEQ activates all seeds at $t = 1$)

\mathcal{E}_2 : \exists an active border node $v_j \in B(u, t)$ in G . (\mathcal{E}_2 implies there exist $\tau < t' \leq t$ such that the border node $X_{j,\tau}$ in \mathcal{H} is active and GETSEQ will activate node $X_{u,t'}$ and u activates by time t .)

We now use the relationship between the capacity of the border nodes and the demand of the (u, t) flow problem (namely, $\sum_{\theta(u) < \tau \leq t} x_{u,\tau}$) to prove Part 1 of (5.6). Given our

observation above we have:

$$\begin{aligned}
\Pr[T(v) \leq t] &= \Pr[\mathcal{E}_1 \vee \mathcal{E}_2] \geq 1 - \min\{\Pr[\neg\mathcal{E}_1], \Pr[\neg\mathcal{E}_2]\} \\
&\geq 1 - \min\left\{\Pr[\neg\mathcal{E}_1], 1 - 2(1 + \epsilon)\left(\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}\right)\right\} \quad (\text{Lemma 5.3.7}) \\
&\geq 1 - \min\left\{1 - 2(1 + \epsilon)\sum_{\tau < \theta(u)} x_{u,\tau}, 1 - 2(1 + \epsilon)\left(\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}\right)\right\} \quad (\text{Since } \alpha \geq 2(1 + \epsilon)) \\
&\geq \max\left\{2(1 + \epsilon)\sum_{\tau < \theta(u)} x_{u,\tau}, 2(1 + \epsilon)\left(\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}\right)\right\} \geq (1 + \epsilon)\sum_{\tau \leq t} x_{u,\tau}
\end{aligned}$$

which completes the proof, modulo Lemma 5.3.7 used in the second inequality. Lemma 5.3.7 applies the border node Lemma 5.3.6 to relate the probability that at least one border node is in the seedset (*i.e.* $\Pr[\neg\mathcal{E}_2]$) with the demand of the (u, t) -flow problem (*i.e.* $\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$). Specifically:

Lemma 5.3.7. *For every $u \in V$ and $t \in [n]$ where $\sum_{\tau \leq t} x_{u,\tau} \leq \frac{1}{12(1+\epsilon)}$ we have*

$$\Pr[\neg\mathcal{E}_2] = 1 - \prod_{v_j \in B(u,t)} (1 - p_j) \geq 2(1 + \epsilon) \sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$$

Proof of Lemma 5.3.7. We shall find a non-negative sequence p'_j ($v_j \in B(u, t)$) such that

- Condition 1: $\prod_{v_j \in B(u,t)} (1 - p_j) \leq \prod_{v_j \in B(u,t)} (1 - p'_j)$
- Condition 2: $\sum_{j \in B(u,t)} p'_j = 4(1 + \epsilon) \sum_{\theta(u) \leq \tau \leq t} x_{u,\tau}$.

When both conditions hold, we can bound $\prod_{v_j \in B(u,t)} (1 - p_j)$ by $\prod_{v_j \in B(u,t)} (1 - p'_j)$, which can then be approximated by its first-order terms. We use existential arguments to find the sequences p'_j (for each $v_j \in B(u, t)$): We start by recalling that $p_j = \min\{1, \alpha\omega_j\}$ and $\alpha > 4(1 + \epsilon)$. It follows that when $\omega_j \geq \frac{1}{4(1+\epsilon)}$ for some $v_j \in B(u, t)$, the $p_j = 1$ and the lemma trivially holds. Thus, we may assume that $4(1 + \epsilon)\omega_j \leq 1$

for all $v_j \in B(u, t)$, and we can write

$$\sum_{v_j \in B(u, t)} p_j \geq 4(1 + \epsilon) \sum_{v_j \in B(u, t)} \omega_j \geq 4(1 + \epsilon) \sum_{\theta(u) \leq \tau \leq t} x_{u, \tau}.$$

where the second inequality uses Lemma 5.3.6.

We now know that there exists a sequence p'_j such that $p_j \geq p'_j$ and $\sum_{j \in B(u, t)} p'_j = 4(1 + \epsilon) \sum_{\theta(u) \leq \tau \leq t} x_{u, \tau}$, which meets Condition 1 and Condition 2. It follows that $\prod_{v_j \in B(u, t)} (1 - p_j) \leq \prod_{v_j \in B(u, t)} (1 - p'_j)$, and we may complete the proof with the following first-order approximation:

Lemma 5.3.8 (First order approximation). *Let x_1, x_2, \dots, x_k be real positive values such that $\sum_{i \leq k} x_i \leq 1$. Then*

$$\prod_{i \leq k} (1 - x_i) \leq 1 - \frac{1}{2} \left(\sum_{i \leq k} x_i \right).$$

Proof of Lemma 5.3.8 (First order approximation). Let x_1, x_2, \dots, x_k be real positive values such that

$\sum_{i \leq k} x_i \leq 1$. Notice that for any $0 \leq x \leq 1$, we have $(1 - x) \leq e^{-x}$. Let $s \triangleq \sum_{i \leq k} x_i$.

We have

$$\prod_{i \leq k} (1 - x_i) \leq \prod_{i \leq k} \exp(-x_i) = \exp\left(-\sum_{i \leq k} x_i\right) = \exp(-s) \leq 1 - s + \frac{s^2}{2} \leq 1 - s\left(1 - \frac{1}{2}\right) = 1 - \frac{s}{2}.$$

□

When we substitute the x_i 's in Lemma 5.3.8 with p'_j s, and use the fact that

$$\sum_{j \in B(u, t)} p'_j = 4(1 + \epsilon) \sum_{\theta(u) \leq \tau \leq t} x_{u, \tau} \leq 4(1 + \epsilon) \frac{1}{12(1 + \epsilon)} = \frac{1}{3} < 1.$$

we complete the proof because $\prod_{v_j \in B(u, t)} (1 - p'_j) \leq 1 - \frac{1}{2} \cdot 4(1 + \epsilon) \sum_{\theta(v_j) \leq \tau \leq t} x_{u, \tau}$

We now move to the second part of Proposition 5.3.1. , *i.e.* we consider a pair (u, t) such that $\sum_{\tau \leq t} x_{u,\tau} \geq \frac{1}{12(1+\epsilon)}$. Let us consider two cases.

Case 1. $\sum_{\tau \leq \min\{\theta(u)-1, t\}} x_{u,\tau} \geq \frac{1}{24(1+\epsilon)}$. In this case, $p_u = 1$ and u is always selected as a seed. Thus, $\Pr[T(u) \leq t] = 1$.

Case 2. $\sum_{\tau \leq \min\{\theta(u)-1, t\}} x_{u,\tau} < \frac{1}{24(1+\epsilon)}$ In this case, we can see that $\sum_{\theta(u) \leq \tau \leq t} x_{u,\tau} \geq \frac{1}{24(1+\epsilon)}$. Therefore, we use the border node Lemma 5.3.6 to get

$$\sum_{v_j \in B(u,t)} \omega_j \geq \frac{1}{24(1+\epsilon)}. \quad (5.7)$$

Now, recall that $\Pr[T(u) \leq t] \geq \Pr[\mathcal{E}_2] = 1 - \prod_{v_j \in B(u,t)} (1 - p_j)$. Therefore, it suffices to prove that $\prod_{v_j \in B(u,t)} (1 - p_j) \leq \frac{\epsilon}{4n^2}$.

At this point, our analysis deviates from the analysis for the first part of Proposition 5.3.1. There, the p_j values were small enough to allow $\prod_{v_j \in B(u,t)} (1 - p_j)$ to be approximated using only first-order terms. Here, we are dealing with the case where p_j 's are large. Thus, $\prod_{v_j \in B(u,t)} (1 - p_j)$ decays exponentially, and it is more appropriate to approximate it using exponential functions. By using (5.7) and the following approximation Lemma 5.3.9 (with λ as α) we can see that indeed $\prod_{v_j \in B(u,t)} (1 - p_j) \leq \frac{\epsilon}{4n^2}$, which completes the proof.

Lemma 5.3.9. *Let ϵ be an arbitrary constant. Let x_1, \dots, x_k be numbers between $[0, 1]$ such that $\sum_{i \leq k} x_i = s$, where $s \geq \frac{1}{24(1+\epsilon)}$. Let $\lambda = 24(1+\epsilon) \ln(\frac{4n^2}{\epsilon})$ and $p_i = \min\{\lambda x_i, 1\}$. It follows that*

$$\prod_{i \leq n} (1 - \min\{\lambda x_i, 1\}) \leq \frac{\epsilon}{4n^2}.$$

Proof of Lemma 5.3.9. Let us consider two cases over the values of x_i . In the first case, there exists some x_i such that $\lambda x_i \geq 1$. For this case, we have $\prod_{i \leq n} (1 -$

$$\min\{\lambda x_i, 1\} = 0 \leq \frac{\epsilon}{4n^2}.$$

In the second case, where all x_i are less than $1/\lambda$, the quantity $\prod_{i \leq k} (1 - p_i) = \prod_{i \leq k} (1 - \lambda x_i)$ is maximized when $x_1 = x_2 = \dots = x_k = \frac{s}{k}$. In other words,

$$\begin{aligned} \prod_{i \leq k} (1 - \lambda x_i) &\leq \left(1 - \frac{\lambda s}{k}\right)^k \\ &= \left(1 - \frac{\lambda s}{k}\right)^{\frac{k}{\lambda s} \lambda s} \\ &\leq e^{-\lambda s} \\ &\leq \exp\left(-\frac{\lambda}{24(1 + \epsilon)}\right) \\ &= \exp\left(-\ln\left(\frac{4n^2}{\epsilon}\right)\right) = \frac{\epsilon}{4n^2}. \end{aligned}$$

□

This completes the proof for Proposition 5.3.1. □

Asymptotic optimality of our rounding algorithm. We pay a factor of ℓ in our rounding algorithm, because we merge ℓ different $\{S, T\}$ samples to make sure all the thresholds are good. But is this really necessary? In Chapter 5.5.3, show that our rounding algorithm is asymptotically optimal in ℓ , by presenting an $\Omega(\ell)$ integrality gap for the LP of Figure 5.3. Our problem instance is composed of ℓ individual gadgets, where the nodes in gadget i have thresholds chosen from a carefully constructed constant-size set. We can force these gadgets to be “independent”, in that sense that if a single $\{S, T\}$ sample causes one of the thresholds in gadget i to be good, we know that whp no threshold in any other gadget can be good. It follows that merging ℓ different $\{S, T\}$ samples, each ensuring that a single threshold is good, is inevitable.

Improvement to the approximation ratio. Observe that in our rounding procedure we require all T_j 's in each of the sampled pairs $\{S_j, T_j\}$ to be feasible (*i.e.* all nodes have to be active at the end of T_j). This requirement is not necessary because the merged T will be feasible even if only *one* of the T_j is feasible. We remark here that this observation can be exploited to improve the algorithm so that it returns a feasible seedset of size $\alpha \cdot \text{opt} + \beta$, where $\alpha = O(r(\log n + \ell))$ and $\beta = O(r\ell \log n)$.

5.4 Proof of Corollary 5.1.3

Let us consider an arbitrary technology diffusion problem $\Pi = \{G, \theta\}$. Let Π^+ and Π^- be the corresponding diffusion problems defined in Definition 5.1.2. Recall that opt is the optimal solution for Π , opt^+ is the optimal solution for Π^+ and opt^- is the optimal solution for Π^- . Let $\mathbf{P} \triangleq \{\lfloor 1 + \epsilon \rfloor, \lfloor (1 + \epsilon)^2 \rfloor, \dots, \lfloor (1 + \epsilon)^q \rfloor\}$, where $q = \log_{1+\epsilon} n + 1$. We next define a new technology diffusion instance $\Pi' = \{G, \theta'\}$ that uses the same graph and $\theta'(u)$ (for each u) is the smallest number in \mathbf{P} that is larger than $\theta(u)$. Notice that $\theta'(u) \leq (1 + \epsilon)\theta(u)$. Let opt' be the size of optimal seed set for Π' . We can run our approximation algorithm on Π' and get a solution, whose size is at most $O(\log^2 n \cdot r \cdot \text{opt}')$ since the number of thresholds in Π' is $\log n$. Because $\theta'(u) \geq \theta(u)$ for all u , a feasible solution in Π' is also a feasible solution in Π . Thus, the seedset returned by our algorithm is feasible and $\text{opt} \leq \text{opt}'$. Similarly, we can see that $\text{opt}^- \leq \text{opt}$ and $\text{opt}' \leq \text{opt}^+$. Therefore, the seedset size can be expressed as $O(\log^2 n \cdot r \cdot \text{opt} \frac{\text{opt}'}{\text{opt}}) = O(\log^2 n \cdot r \cdot \text{opt} \frac{\text{opt}^+}{\text{opt}^-}) = O(\kappa(\Pi, \epsilon)(\log^2 n) \cdot r \cdot \text{opt})$.

5.5 Lower bounds

We now present the following three lower bounds for the technology diffusion problem.

1. *Computational barrier:* the technology diffusion problem is at least as hard as a Set Cover problem, so that our problem does not admit any $o(\ln|V|)$ -approximation algorithm.
2. *Information theoretic barrier:* in the worst case, the optimal solution with a *connected seedset* could be $\Omega(r)$ times larger than the optimal solution.
3. *Integrality gaps:* The simple IP (Figure 5.1 discussed in Chapter 5.2) has an $\Omega(n)$ integrality gap. The augmented IP (presented in Figure 5.3) has an $\Omega(\ell)$ integrality gap.

5.5.1 Computational barrier

We now prove that the technology diffusion problem is at least as hard as the set cover problem. Let us recall the definition (of the optimization version) of the set cover problem: given a finite universe \mathcal{U} and a family \mathbf{S} of subsets of \mathcal{U} , we are interested in finding the smallest subset \mathbf{T} of \mathbf{S} such that \mathbf{T} is a cover of \mathcal{U} , i.e. $\bigcup_{T \in \mathbf{T}} T = \mathcal{U}$. The set cover cannot be approximated within a factor of $(1 - o(1)) \ln n$ unless NP has quasi-polynomial time algorithm (see [3] and references therein). We have the following lemma.

Lemma 5.5.1. *Given an α -approximation algorithm for the technology diffusion problem with constant number of threshold values $\theta \geq 2$, and constant graph diameter $r \geq 3$, we can obtain an $O(\alpha)$ -approximation algorithm for the set cover problem.*

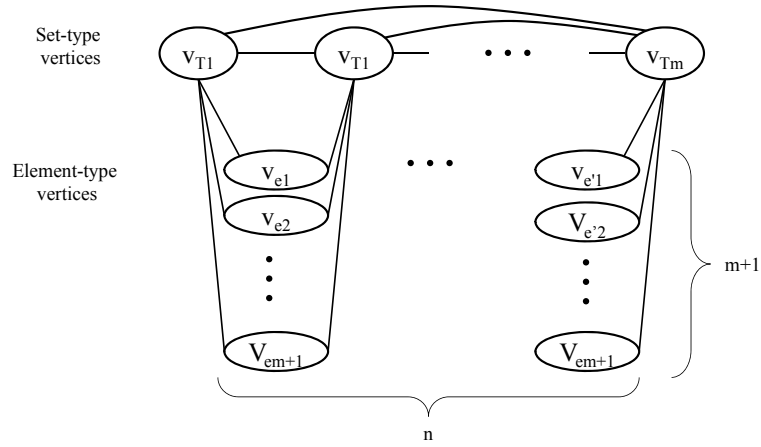


Figure 5.4: Reduction.

Moreover, the reduction holds even if the seedset in the technology diffusion problem is required to be connected.

Thus, we can see that there is no $c \ln n$ approximation algorithm (for some constant c) for the technology diffusion problem.

Proof of Lemma 5.5.1. Let us consider an arbitrary set cover instance $(\mathcal{U}, \mathbf{T})$, where $m = |\mathbf{T}|$ is the number of sets in \mathbf{T} .

The reduction. We construct a technology diffusion problem as described below, and illustrated in Figure 5.4:

- The vertex set consists of the following types of vertices:
 1. The *set type*: for each $T \in \mathbf{T}$, we shall construct a node v_T in the technology network.
 2. The *element type*: for each $e \in \mathcal{U}$, we shall construct $m + 1$ nodes $v_{e,1}, v_{e,2}, \dots, v_{e,m+1}$.

- The edge set consists of the following edges:
 1. For each $T \in \mathbf{T}$ and $e \in T$, we add the edges $\{v_T, v_{e,1}\}, \{v_T, v_{e,2}\}, \dots, \{v_T, v_{e,m+1}\}$.
 2. The set type vertices are connected as a clique. (For each $T \neq T' \in \mathbf{T}$, we add the edge $\{u_T, u_{T'}\}$).
- The thresholds $\theta(\cdot)$ are set as follows,
 1. For any $e \in \mathcal{U}$ and $i \leq m + 1$, we set $\theta(v_{e,i}) = 2$.
 2. For every $T \in \mathbf{T}$, we set $\theta(v_T) = (m + 1)n + 1$.

Properties of the reduction. Notice that our technology diffusion problem has only two types of threshold values. Furthermore, the diameter of the graph we form is exactly 3 hops (in terms of edges); the maximum distance in this graph is from one $v_{e,i}$ node to another. Finally, we show below that the seedset must consist of set-type vertices. Since these vertices form a clique, it follows that there exists an optimal seedset that is connected.

Correctness. To conclude that the size of the optimal seed set is the same as the size of the optimal cover (which also means that our reduction is approximation-preserving), we establish the following:

Item 1. For any feasible cover \mathbf{S} in the set cover problem, the corresponding seed set $\{v_S : S \in \mathbf{S}\}$ is a feasible solution for the technology diffusion problem.

Item 2. Any feasible seedset in the technology diffusion problem *that only consists of set-type vertices* corresponds to a feasible cover in the set cover problem.

Item 3. Given a feasible seedset that consists of *element type* vertices, there is a feasible seedset of equal or smaller size that consists only of *set type* vertices. Since the set type vertices form a clique, we have that the optimal solution for the technology diffusion problem is also a connected one.

Item 1. To show the first item, we simply walk through the activation process: When \mathcal{S} is a cover, let the seedset be v_{T_i} for all $T_i \in \mathcal{S}$. Notice that this seedset is *connected*. Upon activating the seedset, the vertices $u_{e,i}$ for all $e \in \mathcal{U}$ and $i \leq m+1$ are activated because they are connected to at least one active seed. Now, there are $(m+1)n$ active nodes, so the rest of the *set type* vertices are activated.

Item 2. To show the second item, we consider an arbitrary seedset that only consists of the *set type* vertices: $U = \{v_{T_1}, v_{T_2}, \dots, v_{T_k}\}$, where $T_1, \dots, T_k \in \mathbf{T}$. We shall show that if T_1, \dots, T_k is not a cover, then the seed set cannot be feasible (*i.e.* some nodes will remain inactive in the technology diffusion problem).

Let $e \in \mathcal{U} / (\cup_{j \leq k} T_j)$ be an element that is not covered by the sets in $\{T_1, \dots, T_k\}$. Let us consider the nodes $v_{e,1}, v_{e,2}, \dots, v_{e,m+1}$, and node v_T for each $T \notin \{T_1, \dots, T_k\}$ in the technology diffusion problem. We claim that none of these vertices will be activated with seedset U . Suppose, for the sake of contradiction, that one or more of these vertices are activated, and consider the first activated vertex among them. There are two cases:

Case 1. v_T ($T \notin \mathbf{T}$) is activated first. This is impossible: when $v_{e,i}$ ($i \leq m+1$) are not activated, the number of activated nodes is at most $(n-1)(m+1)+m < (m+1)n$.

Case 2. $v_{e,i}$ ($i \leq m+1$) is activated first. This is impossible because $v_{e,i}$ is only

connected with v_T , where $T \notin \{T_1, \dots, T_k\}$ and none these *set type* vertices are activated.

Item 3. Finally, we move onto the third item. Let us consider a feasible seedset F that does not consist of only *set type* vertices. We show that we can easily remove the *element type* vertices in F : let $v_{e,i}$ be an arbitrary vertex in F . Then we can remove $v_{e,i}$ from F and add an v_T to F such that $e \in T$. This does not increase the cardinality of F . Furthermore, $v_{e,i}$ would still be activated, which implies that the updated F is still be a feasible seed set. \square

5.5.2 Information theoretic bound

Lemma 5.5.2. *For any fixed integer r , there exists an instance of technology diffusion problem $\{G, \theta\}$ such that (a) the diameter of G is $\Theta(r)$, and (b) the optimal connected seedset is at least $\Omega(r)$ larger than the optimal seedset.*

Proof of Lemma 5.5.2. Let $r > 0$ be an arbitrary integer. We define graph G_r as follows (see Figure 5.5):

- The vertex set is $\{v_1, \dots, v_{2r+1}\}$.
- The edge set is $\{\{v_i, v_{i+1}\} : 1 \leq i < 2r + 1\}$.

The threshold function shall be defined as follows,

- $\theta(v_1) = \theta(v_{2r+1}) = 2$ and $\theta(v_{r+1}) = 2r + 1$.
- For $1 < i \leq r$, $\theta(v_i) = i$.
- For $r + 2 \leq i < 2r + 1$, $\theta(v_i) = 2r + 2 - i$.

It is straightforward to see that the diameter of the graph is $2r = \Theta(r)$. It remains to verify that the optimal connected solution is $\Theta(r)$ times larger than the optimal solution.

It's easy to see that $\{v_1, v_{2r+1}\}$ is a feasible seedset and therefore, the size of the optimal seed set is $O(1)$. We next show that any feasible connected set has size $\Omega(r)$.

Since the seedset must be connected, wlog we can assume that the seedset is $\{v_i, v_{i+1}, \dots, v_j\}$ and by symmetry $i \leq r + 1$. When $j < r + 1$, node v_{r+1} will never activate (because v_{r+1} has threshold $2r + 1$, it only activates when all other nodes are active, but in this case all r nodes to the right of v_{r+1} are inactive). It follows that a feasible seedset requires $j \geq r + 1$.

When $i = 1$, the size of the seedset is $\Theta(r)$ and the lemma follows. So, we need only consider the case where $i > 1$: symmetry allows us to assume wlog that $r + 1 - i \geq j - (r + 1)$ i.e. $\theta(v_{j+1}) \geq \theta(v_{i-1})$. Therefore, since we have $j - i + 1$ nodes in the seedset, a necessary condition for this seedset to be feasible is thus $j - i + 1 \geq i - 2$. Using the fact that $j \geq r + 1$, we get $i \leq r/2 + 2$ and $j - i = \Omega(r)$, which completes our proof. \square

One drawback of this construction is that $\ell = \Theta(n)$. We may modify $\theta(\cdot)$ so that $\ell = \tilde{O}(1)$ (thus ensuring that our lower bound depends on graph diameter r , rather than the number of thresholds ℓ):

- When $i \leq n$, set $\theta(u_i) = \max\{2^{\lfloor \log_2 i \rfloor}, 2\}$,
- when $i = n + 1$, set $\theta(u_i) = 2n + 1$, and
- When $i > n$, set $\theta(u_i) = \max\{2^{\lfloor \log_2 (2n+2-i) \rfloor}, 2\}$.

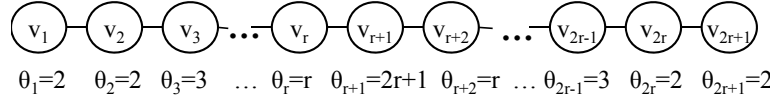


Figure 5.5: An instance of the technology diffusion problem for the proof of Lemma 5.5.2.

One can use similar arguments to show that the size of the optimal seedset is $O(1)$ while the size of the optimal connected seedset is $\Theta(r)$.

5.5.3 Integrality gap

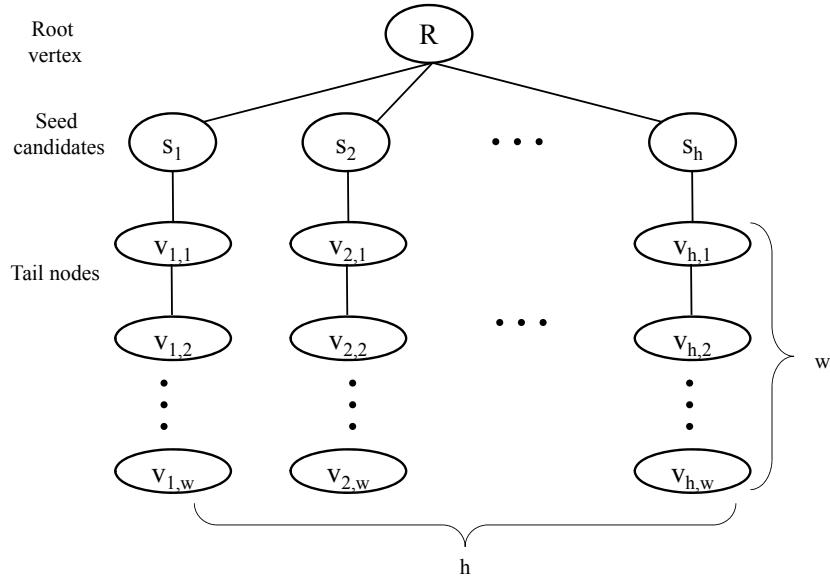
Integrality gap for the simple IP of Figure 5.1

We construct a problem instance with $\ell = O(1)$ where the solution returned by the simple IP of Figure 5.1 is $O(1)$, while the optimal seedset has size $\Theta(n)$, implying an integrality gap that is polynomial in $n = |V|$.

The problem instance. We let w and h be parameters of the problem instance $\{G, \theta\}$. These parameters control the shape of the graph G and the size of the integrality gap. We will decide the parameters at the end to maximize the integrality gap. The graph G (see Figure 5.6) has a node set of size $n = wh + h + 1$ that consists of the following nodes:

- The root node R .
- The “seed candidates” $\{s_1, \dots, s_h\}$.
- The “tail nodes” $v_{i,j}$ for $i \leq h$ and $j \leq w$.

The edge set consists of the following two types of edges:

Figure 5.6: The graph for the hard instance with $\Omega(h)$ integrality gap

- all the “seed candidates” s_i ($i \in [h]$) are connected with the root R .
- for any specific $i \in [h]$, the nodes $s_i, v_{i,1}, v_{i,2}, \dots, v_{i,w}$ form a chain. In other words, $\{s_i, v_{i,1}\} \in E$ and $\{v_{i,j}, v_{i,j+1}\} \in E$ for $1 \leq j \leq w - 1$.

Hereafter, we shall refer to the chain $s_i, v_{i,1}, \dots, v_{i,w}$ as the i -th tail of the graph. The threshold function θ is specified as follows:

- $\theta(R) = n$.
- for any s_i we have $\theta(s_i) = n - h + 2$.
- for any $v_{i,j}$ we have $\theta(v_{i,j}) = 2$.

To exhibit the integrality gap, we shall first construct a feasible fractional solution of constant size, and then show that the optimal integral solution gives rise to a seedset of size $\Theta(h)$.

The fractional solution. Table 5.1 describes a feasible fractional solution of constant size. We now walk through this solution. In the solution, we group the rows in the following way:

- The first row corresponds with the root node.
- The rest of the rows are grouped by “stripes”. A stripe consists of a seed candidate and its corresponding tail. For example, the first stripe consists of the rows for $s_1, v_{1,1}, \dots, v_{1,w}$.

We shall also divide the columns into two parts. The first part is the “false propagation” stage, consisting of $2 + wh$ columns where we use a small fractional seed set to activate the tail nodes. The second part is the “completion stage” consisting of $h - 1$ rows, where we fill in the residual mass of the nodes so that the permutation constraints are met.

Variable assignments in the fractional solution. We now describe the assignments in Table 5.1.

- $x_{R,1} = 1$, i.e. the root is first activated.
- Let $\epsilon \triangleq 1/h$. For each stripe $\{s_i, v_{i,1}, \dots, v_{i,w}\}$, we assign values in the false propagation region as follows: $x_{s_i,2} = \epsilon$ and $x_{v_{i,j},j+2+kw} = \epsilon$ for all $j \leq w$ and $0 \leq k < h$. The rest of the variables in this region are set to 0. This assignment exhibits a periodic pattern, so that mass can circulate back and forth along a tail until all nodes in the tail are activated at the end of the false propagation stage. (Refer to the underlined values in the first stripe of Table 5.1).

Table 5.1: A fractional solution for the simple IP formulation

		First cycle						Second cycle				$h - 2$ other cycles				Completion stage			
1st stripe	R	1	0	0	0	0	...	0	0	0	...	0	...	0	0	0	...	0	
	s_1	0	ϵ	0	0	0	...	0	0	0	...	0	...	$1 - \epsilon$	0	0	...	0	
	$v_{1,1}$	0	0	$\underline{\epsilon}$	0	0	...	0	$\underline{\epsilon}$	0	...	0	...	0	0	0	...	0	
	$v_{1,2}$	0	0	0	$\underline{\epsilon}$	0	...	0	0	$\underline{\epsilon}$...	0	...	0	0	0	...	0	
	$v_{1,3}$	0	0	0	0	$\underline{\epsilon}$...	0	0	0	$\underline{\epsilon}$...	0	...	0	0	...	0	
	\vdots						\ddots				\ddots		\ddots				...		
$v_{1,w}$	0	0	0	0	0	...	$\underline{\epsilon}$	0	0	0	...	$\underline{\epsilon}$...	0	0	0	...	0	
2nd stripe	s_2	0	ϵ	0	0	0	...	0	0	0	...	0	...	ϵ	$1 - 2\epsilon$	0	...	0	
	$v_{2,1}$	0	0	ϵ	0	0	...	0	ϵ	0	...	0	...	0	0	0	...	0	
	$v_{2,2}$	0	0	0	ϵ	0	...	0	0	ϵ	...	0	...	0	0	0	...	0	
	$v_{2,3}$	0	0	0	0	ϵ	...	0	0	0	ϵ	...	0	...	0	0	...	0	
	\vdots						\ddots				\ddots		\ddots				...		
	$v_{2,w}$	0	0	0	0	0	...	ϵ	0	0	0	...	ϵ	...	0	0	0	...	0
3rd stripe	s_3	0	ϵ	0	0	0	...	0	0	0	...	0	...	0	2ϵ	$1 - 3\epsilon$...	0	
	$v_{3,1}$	0	0	ϵ	0	0	...	0	ϵ	0	...	0	...	0	0	0	...	0	
	$v_{3,2}$	0	0	0	ϵ	0	...	0	0	ϵ	...	0	...	0	0	0	...	0	
	$v_{3,3}$	0	0	0	0	ϵ	...	0	0	0	ϵ	...	0	...	0	0	...	0	
	\vdots						\ddots				\ddots		\ddots				...		
	$v_{3,w}$	0	0	0	0	0	...	ϵ	0	0	0	...	ϵ	...	0	0	0	...	0
h-th stripe	\vdots	\vdots	\vdots			\vdots			\vdots	\vdots							\vdots		
	s_h	0	ϵ	0	0	0	...	0	0	0	...	0	...	0	0	0	...	$(h - 1)\epsilon$	
	$v_{h,1}$	0	0	ϵ	0	0	...	0	ϵ	0	...	0	...	0	0	0	...	0	
	$v_{h,2}$	0	0	0	ϵ	0	...	0	0	ϵ	...	0	...	0	0	0	...	0	
	$v_{h,3}$	0	0	0	0	ϵ	...	0	0	0	ϵ	...	0	...	0	0	...	0	
	\vdots						\ddots				\ddots		\ddots				...		
$v_{h,w}$	0	0	0	0	0	...	ϵ	0	0	0	...	ϵ	...	0	0	0	...	0	

- Finally, we fill in the variables in the completion stage so that the permutation constraints are met. Notice that at time $n-h+2$, only the rows that correspond with the seed candidates s_i ($i \leq h$) do not sum up to 1. We use the columns in the completion stage to fill in the extra mass using a “greedy” approach. In other words, at the column $n-h+2$, we first fill in the unused mass (namely $1-\epsilon$) from s_1 . Then we fill in the unused mass from s_2 as much as possible, subject to the constraint that the column sums to 1 (namely ϵ). Next, we move to the next column (the $n-h+3$ -rd column). Then we fill in the mass from s_2 and as much mass as possible from s_3 to this column. This process continues until all mass from s_i ($i \leq h$) is filled.

The fractional solution is feasible. Next, we argue that such assignments are feasible. Since, we satisfied the permutation constraints by construction (Table 5.1), we only argue that the connectivity constraints are met.

- We need to start thinking about connectivity when $t = 2$. At this time step, the connectivity constraints are met because all the seed candidates are connected to the root R , which is activated at time $t = 1$.
- Next, we argue that the connectivity constraints are met in the propagation stage. Let us consider the first cycle in the propagation stage. In the first time step of the first cycle, an ϵ -fraction of mass is activated at $v_{i,1}$ for all $i \leq h$. Since $v_{i,1}$ is connected with s_i , and an ϵ portion of s_i is active prior to the beginning of the 1st cycle, the connectivity constraint is met for this step. For the rest of the timesteps of the first cycle, note that by the time we assign ϵ to the node $v_{i,j}$, an ϵ portion of mass is already activated at $v_{i,j-1}$. Since $\{v_{i,j-1}, v_{i,j}\} \in E$

for all $j < w$, the connectivity constraints are met for the entire first cycle. The argument for the remaining cycles proceeds in a similar manner.

- Finally, showing the connectivity holds in the completion stage is trivial: this follows because only seed candidates activate at this stage, and seed candidates are all connected to the root which has been fully activated since $t = 1$.

Hence, we can conclude that the fractional solution in Table 5.1 is feasible.

The integral solution. To prove that the optimal integral solution is a seedset of size $O(h)$, we show that any seedset of size less than $\frac{h}{5}$ will fail to activate all the nodes in the graph. Here, the constant $\frac{1}{5}$ is chosen rather arbitrarily and is not optimized.

First, we notice that for any feasible set S that contains one or more tail nodes, we can transform it into a feasible set S' such that (a) $|S'| \leq |S|$ and (b) no tail nodes are in S' . To construct the new seedset, replace each tail node $v_{i,j}$ in S by its parent seed candidate s_i . Since the activation of s_i always causes the activation of $v_{i,j}$ for any j , it follows that S' is a feasible seedset whenever S is a feasible seedset.

Thus, we may focus on the seedset that contains only R and/or seed candidates. Wlog, we may assume the seed set is a subset in $U = \{R, s_1, \dots, s_{\frac{h}{5}}\}$. Next, we argue that the seedset U fails to activate all the nodes in the graph. First, we can see that all the tails $v_{i,j}$ ($i \leq \frac{h}{5}$ and $j \leq w$) with parent seed candidates in U will be active. After they are activated, the total number of activated nodes will be $\frac{h}{5} + 1 + \frac{wh}{5}$. Now we argue that no other nodes are active because (a) all seed candidates s_i ($i > \frac{h}{5}$) that are not in U cannot be activated since the following holds

$$\left(\frac{h}{5} + 1 + \frac{wh}{5}\right) + 1 < \theta(s_i) = n - h + 2 \quad (5.8)$$

for sufficiently large constant w and sufficiently large n , and (b) all tail nodes $v_{i,j}$ ($i > \frac{h}{5}$ and $j \leq w$) cannot be activated until their parent seed candidate is active.

A $\Theta(n)$ integrality gap. We can conclude that the integral solution has a seedset size of $O(h)$ while the fractional solution is $O(1)$. When we set w be a sufficiently large constant and $h = \Theta(n)$ (we only need to ensure that (5.8) holds), our integrality gap is $\Theta(n)$.

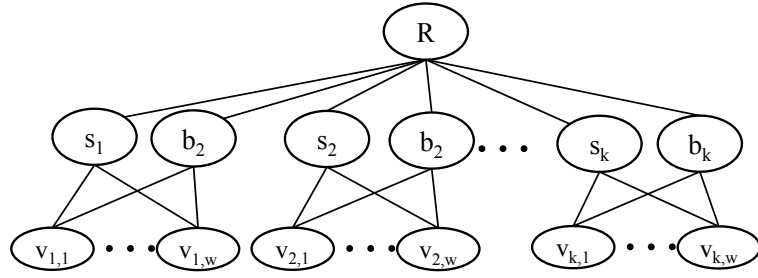
Integrality gap for the augmented IP of Figure 5.3.

We next prove the following theorem.

Theorem 5.5.3. *Consider the augmented linear program of Figure 5.3. For any sufficiently large n and any $\ell \leq cn^{1/3}$, where c is a suitable constant, there exists a problem instance with an $\Omega(\ell)$ integrality gap.*

The problem instance. To simplify the exposition, we will assume that our problem instance $\{G, \theta\}$ is such that our graph G has $|V| = n$ nodes, where $n - 1$ is a multiple of ℓ , and the range of θ is $2\ell + 2$ different threshold values. We shall let w be the integer such that $(w + 2)\ell + 1 = n$, and let $\epsilon \triangleq 1/\ell$. Our graph G is described as follows (See Figure 5.7):

- the node set consists of the following:
 - The root vertex R .
 - The set of “seed candidate” $\{s_1, \dots, s_\ell\}$.
 - The set of “blockers” $\{b_1, \dots, b_\ell\}$.
 - The set of “tails” $v_{i,j}$, where $i \leq \ell$ and $j \leq w$.

Figure 5.7: The graph for the hard instance with $\Omega(k)$ integrality gap

- The edge set consists of the following three types of edges
 - There is an edge between the root and any seed candidate, i.e. $\{R, s_i\} \in E$ for all $i \leq \ell$.
 - There is an edge between the root and any blocker, i.e. $\{R, b_i\} \in E$ for all $i \leq \ell$.
 - For any i, j , we have $\{s_i, v_{i,j}\} \in E$ and $\{b_i, v_{i,j}\} \in E$.

In what follows, we shall also refer to the subgraph induced by $s_i, b_i, v_{i,1}, \dots, v_{i,w}$ as the i -th gadget of the graph. We set the threshold function θ as follows:

- $\theta(R) = n$.
- $\theta(s_i) = (w + 1)\ell + 3$.
- $\theta(v_{i,j}) = (i - 1)(w + 1) + 3$.
- $\theta(b_i) = (i - 1)(w + 1) + w/\ell + 2$.

In what follows, we shall first show that a feasible solution of size $O(1)$ exists for the relaxed LP. Then we shall show that the optimal integral solution is a seedset of size $\Omega(\ell)$.

The fractional solution. We now construct feasible fractional solution of size $O(1)$. See the table in Figure 5.2. The intuition behind our construction is to activate the root R at the first time step, i.e. $x_{R,1} = 1$, and then activate an ϵ -portion of each seed candidate in the second time step. This total of $1 + \ell\epsilon = 2$ mass will be the size of the entire (fractional) seedset. We will make sure that the rest of the node's mass will activate after their thresholds, and will therefore not contribute to the size of the fraction solution.

We divide our construction into two time stages. The first is the “false propagation” stage, where all the nodes except for the seed candidates will be fully activated. The second is the “completion stage” where the remaining inactivated mass from the seed candidates will be activated. Next, we describe each of these two stages in detail.

False propagation stage. The false propagation stage consists of $(w + 1) \times \ell$ time steps, divided into ℓ blocks, each consisting of $(w + 1)$ time steps. Notice that the thresholds of the seed candidates $\theta(s_i) \forall i$ occur exactly after the false propagation stage ends. During i -th block of the false propagation stage, the blocker b_i and tail nodes $v_{i,j} \forall j \in [w]$ in i -th gadget will be fully activated. Since there are exactly $(w + 1)$ such nodes, the i -th block is a $(w + 1) \times (w + 1)$ matrix, the only non-zero variables in the i -th block will be those of the blocker and tail nodes in the i -th gadget. These variables are expressed as the sub-matrices M in Table 5.2.

The variable assignments in M . We next describe the variable assignments in M for the i -th block, as shown in Table 5.3. Our variable assignments will keep the invariance that before the i -th block, all nodes in the k -th gadget, *except* for the seed candidate nodes, are fully activated for every $k \leq i - 1$.

Table 5.2: Feasible fractional assignments for the flow based linear program.

			Propagation Stage				Completion Stage				
R	1	0	0 ...	0	0 ...	0	0	0 ...	0	
s_1	0	ϵ	0 ...	0 ...		0 ...	$1 - \epsilon$	0	0 ...	0	
s_2	0	ϵ	0 ...	0 ...		0 ...	ϵ	$1 - 2\epsilon$	0 ...	0	
\vdots	\vdots	\vdots	\vdots	...	\vdots		
s_ℓ	0	ϵ	0 ...	0 ...		0 ...	0	0	0 ...	$(\ell - 1)\epsilon$	
b_1	0	0	M	0	...	0	0	0	0 ...	0	
$v_{1,1}$	0	0					0	0	0 ...	0	
\vdots	\vdots	\vdots						...			
$v_{1,w}$	0	0					0	0	0 ...	0	
b_2	0	0	0	M	...	0	0	0	0 ...	0	
$v_{2,1}$	0	0					0	0	0 ...	0	
\vdots	\vdots	\vdots						...			
$v_{2,w}$	0	0					0	0	0 ...	0	
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots		
b_ℓ	0	0	0	0	...	M	0	0	0 ...	0	
$v_{\ell,1}$	0	0					0	0	0 ...	0	
\vdots	\vdots	\vdots						...			
$v_{\ell,w}$	0	0					0	0	0 ...	0	

The i -th block begins at the $3 + (i - 1)(w + 1)$ -th time step and ends at the $2 + i(w + 1)$ -th time step. The assignments in M are divided into multiple *cycles*, each of which spans $w/\ell = \epsilon w$ time steps. Notice that the i -th block will contain in total ℓ cycles, and one extra time step that does not belong to any cycle. This extra time step will be inserted between the end of the first cycle and the beginning of the second cycle and will be used to activate the blocker b_i , *i.e.* $x_{b_i, 3+(i-1)(w+1)+w/\ell} = 1$.

In each cycle, every node's mass needs to be incremented by ϵ . We do this using a greedy construction, incrementing the mass of ℓ tail nodes by ϵ in each timestep of a given cycle, so that the column constraints are met for this cycle. For this reason, we need $\frac{w}{\ell}$ time steps to fully activate all the tail nodes; this follows because there are in total ℓ cycles, so the sum of the active portion of any tail node $v_{i,j}$ in the i -th block is $\epsilon \cdot \ell = 1$, so that $v_{i,j}$ is completely activated.

Feasibility of the assignments. We show why our variable assignments at the propagation stage are feasible, and do not increase the mass of the fractional seedset. Our analysis is based on induction. Recall our invariance that prior to the start of the i -th block, all blockers and tail nodes in the j -th gadget ($k \leq i - 1$) are fully activated. We shall show that if the invariance holds up to the $(i - 1)$ st block, the variable assignments in the i -th block are feasible (and do not introduce any mass to the seedset). Suppose that the invariance holds up to the $(i - 1)$ -st block. Then we have that:

- the seedset does not increase, because mass for each node is assigned after its corresponding threshold. (This follows since the i -th block starts at timestep $(w + 1)(i - 1) + 3$ and the tail nodes have $\theta(v_{i,j}) = (w + 1)(i - 1) + 3$. Similarly,

the blocker b_i is activated at time $(w + 1)(i - 1) + 3 + w/\ell$ while $\theta(b_i) = (w + 1)(i - 1) + 3 + w/\ell$.

- the flow constraints are met. During the first cycle, (timesteps $(w + 1)(i - 1) + 3$ to $(w + 1)(i - 1) + w/\ell + 2$), we may push a flow of size ϵ to any tail node $v_{i,j}$ through the path $R-s_i-v_{i,j}$ since the seed candidate s_i has is an ϵ -portion active. Next, at timestep $(w + 1)(i - 1) + w/\ell + 3$ the blocker b_i must receive a unit flow; this is feasible since b_i is directly connected to the root that is fully active at $t = 1$. Finally, during the remaining cycles (timesteps $(w + 1)(i - 1) + w/\ell + 4$ to $(w + 1)i + 2$) we may continue to fully activate the tail nodes $v_{i,j}$ by pushing a up to a unit of flow through the path $R-b_i-v_{i,j}$.

Thus, we may conclude that the assignments at the i -th block are feasible and will not increase the size of the seedset, which further implies that the invariance also holds for the i -th block.

Completion stage. We now describe the assignments for the completion stage. The completion stage starts at time $t_2 \triangleq 3 + \ell(w + 1)$. Since $t_2 \geq \theta(s_i)$, activating the seed candidates in this stage does not increase the side of the seedset. Note further that the only rows that do not sum to 1 correspond to the seed candidates. We again take a greedy approach to fill in the residual mass from the seed candidates (similar to that used in the completion stage of the integrality gap presented in Chapter 5.5.3). At column t_2 , we let $x_{s_1, t_2} = 1 - \epsilon$ and $x_{s_2, t_2} = \epsilon$, filling in the unused row mass from s_1 , and filling in the unused row mass of s_2 as much as possible subject to the column constraint of column t_2 . We repeat this process for each of the $\ell - 1$ remaining columns, until all mass from the seed candidates is used up.

By construction, the completion stage satisfies the row and column permutation constraints, and does not increase the size of the fractional seedset (since all seed candidates activate after their thresholds). Finally, the flow constraints are satisfied since we can push up to a unit of flow to each s_i along its direct connection to R .

The integral solution. To prove that the optimal integral solution is a seedset of size $\Omega(\ell)$, we show that any seedset of size $\frac{\ell}{3}$ fails to activate the whole graph when $w \geq ck^2$ for some suitable constant c . Fix an arbitrary seedset S . Let

$$\mathcal{I} \triangleq \{i : \exists v \in S \text{ s.t. } v \text{ is in the } i\text{-th gadget}\}. \quad (5.9)$$

We shall write $\mathcal{I} = \{i_1, \dots, i_q\}$, where $q \leq \frac{\ell}{3}$. Next, let k be the smallest integer that is not in \mathcal{I} . We proceed to construct a superset S' of S and argue that S' still fails to activate the whole graph. The set S' is constructed as follows:

- Any nodes that are in S are also in S' .
- The root R is in S' .
- Any tail or blocker nodes that are in the first $(k-1)$ -st gadgets are in S' .

We argue that S' will not activate any additional nodes in the graph. For the sake of contradiction, suppose $u \notin S'$ is the first node activated when S' is the seedset. There are two cases:

Case 1. u is in the k -th gadget. The topology of the graph G ensures that any tail node $v_{k,j}$ cannot be activated before s_k or b_k . Therefore, u cannot be $v_{k,j}$. One can see that $\min\{\theta(b_k), \theta(s_k)\} = \theta(b_k) = (k-1)(w+1) + 3 + \epsilon w$. On the other hand, we only have

$$|S'| \leq (k-1)(w+1) + \frac{\ell}{3} + 1.$$

Table 5.3: Matrix \mathbf{M} , the fractional assignments for a block

	First cycle							Second cycle							Third cycle							...		$\frac{w}{\ell}$ -th cycle								
b_i	0	0	0	...	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	...	0	0	0	0	
$v_{i,1}$	ϵ	0	0	...	0	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	
$v_{i,2}$	ϵ	0	0	...	0	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	
\vdots	\vdots			...		\vdots	\vdots	\vdots			...		\vdots	\vdots			...		\vdots	...	\vdots			...		\vdots			...		\vdots	
$v_{i,\ell}$	ϵ	0	0	...	0	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	ϵ	0	0	...	0	0	
$v_{i,\ell+1}$	0	ϵ	0	...	0	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	
$v_{i,\ell+2}$	0	ϵ	0	...	0	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	
\vdots	\vdots			...		\vdots	\vdots	\vdots			...		\vdots	\vdots			...		\vdots	...	\vdots			...		\vdots			...		\vdots	
$v_{i,2\ell}$	0	ϵ	0	...	0	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	0	ϵ	0	...	0	0	
\vdots	\vdots	\vdots		...		\vdots	\vdots	\vdots	\vdots		...		\vdots	\vdots	\vdots		...		\vdots	...	\vdots	\vdots		...		\vdots			...		\vdots	
$v_{i,w-\ell+1}$	0	0	0	...	0	ϵ	0	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ	...	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ
$v_{i,w-\ell+2}$	0	0	0	...	0	ϵ	0	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ	...	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ
\vdots	\vdots			...		\vdots	\vdots	\vdots			...		\vdots	\vdots			...		\vdots	...	\vdots			...		\vdots			...		\vdots	
$v_{i,w}$	0	0	0	...	0	ϵ	0	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ	...	0	0	0	...	0	ϵ	0	0	0	...	0	ϵ

active nodes. Therefore, when $\frac{\ell}{3} \leq \epsilon w$ so that $\ell < cn^{1/3}$ for a sufficiently small constant c , neither b_k nor s_k can be active, and so we have a contradiction.

Case 2. u is not in the k -th gadget. In this case $\theta(u) \geq k(w + 1) + 3$. On the other hand, $|S'| \leq (k - 1)(w + 1) + \frac{\ell}{3} + 1$, so that when $\ell < cn^{1/3}$ for sufficiently large ℓ , the total number of active nodes is less than $\theta(u)$, which is also a contradiction.

The integrality gap. The optimal fractional solution has size $O(1)$, while the optimal integral solution is a seedset of size $\Omega(\ell)$ (when $\ell < cn^{1/3}$ for large enough constant c), so our integrality gap is $\Omega(\ell)$.

5.5.4 Remark on the role of flow constraints in reducing the integrality gap

Finally, we remark the role of flow constraints in reducing the integrality gap from $O(n)$ to $O(\ell)$. From the two gap instances we presented, we can see that there are two types of “bad” mass that can adversarially impact the quality of the linear program:

1. The recirculation of “fake” mass, as discussed in the pathological example of Chapter 5.3.2. We used recirculation of mass to construct the gap instance for the simple IP of Figure 5.1 in Chapter 5.5.3.
2. A chain of fractional mass. Recall that both our gap instances (Chapter 5.5.3 and Chapter 5.5.3), used a seed candidate s_i to connect to a set of $w = 1/\epsilon$ tail nodes $v_{i,1}, \dots, v_{i,w}$ so that when an ϵ -portion of s_i becomes active, the total active fractional mass is $\epsilon \cdot (w + 1) > 1$. Meanwhile, in the integral solution,

we need to activate at least one seed to have a full unit of active mass, which creates a gap of size $1/\epsilon$.

The flow constraints eliminate “bad” mass of the first type (see Chapter 5.3.2), but cannot eliminate the second type. It turns out that if we only have the second type of “bad” mass, the integrality gap becomes $O(\ell)$ instead of $O(n)$.

For ease of exposition, we explain the relationship between the gap and ℓ by referring to the problem instance presented in Chapter 5.5.3. These arguments can also be generalized to other problem instances. Our crucial observation is that the blockers in each of the gadgets have different thresholds. To see why, suppose that two or more gadgets had blockers that did share the same threshold. Observe that if we add a seed candidate from one of these gadgets to the seedset, all the nodes in all these gadgets will become active (because the blockers all have the same threshold). This means that we need to include fewer nodes in seedset for the optimal integral solution, which reduces the size of the integrality gap. To sum up, the idea behind our gap instance is to pad k parallel gadgets together to get a gap of size $\Theta(k)$; for this padding to work we need at least $\Theta(k)$ different threshold values, and so the granularity of the threshold function ℓ scales linearly with the integrality gap.

5.6 Our problem is neither submodular nor super-modular

We wondered about the relationship between the algorithmic properties of our model and the linear threshold model on social networks articulated in [69]. [20]

showed that the problem of selecting an optimal seedset in the linear threshold mode in social networks cannot be approximated within a factor of $O(2^{\log^{1-\epsilon}|V|})$ when the thresholds are deterministic and known to the algorithm. [69] got around this lower bound by assuming that nodes' thresholds are chosen uniformly at random *after* the seedset is selected, and designing an algorithm that chooses the optimal seedset *in expectation*. Their $(1 - 1/e - \epsilon)$ -approximation algorithm relies on the submodularity of the *influence function*, *i.e.* the function $f(S)$ which gives the expected number of nodes that activate given that nodes in S are active.

In this section, we shall show that algorithmic results for submodular and/or supermodular optimization do *not* directly apply to our problem, even if we restrict ourselves to (a) graphs of constant diameter, (b) diffusion problems with a small number of fixed thresholds, or if (c) we choose the thresholds uniformly at random as in [69]. Moreover, we see neither diminishing, nor increasing marginal returns even if we restrict ourselves to (d) connected seedsets.

5.6.1 Fixed threshold case

In this section, we construct two families of technology diffusion instances where the threshold function θ is given as input. Each family will be on a graph of diameter at most 4, and require at most 2 different threshold values, and each will consider connected seedsets. The first family will *fail* to exhibit the submodularity property while the second will *fail* to exhibit supermodularity.

Let $\{G, \theta\}$ be an arbitrary technology diffusion problem. We shall write $f_{G,\theta}(S)$ be the total number of nodes that eventually activate after seedset S activates. When

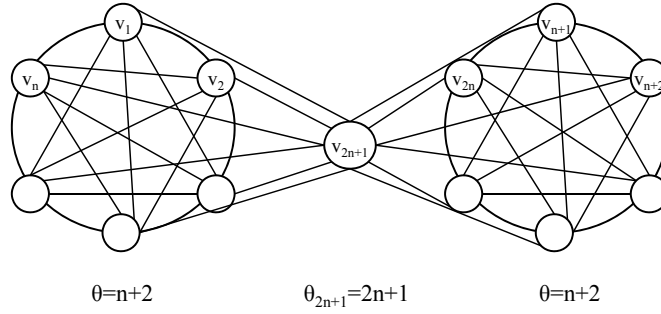


Figure 5.8: An instance of the technology diffusion problem.

G and θ are clear from the context, we simply refer to $f_{G,\theta}(S)$ as $f(S)$.

The influence function is not submodular.

Let n be a sufficiently large integer such that the number of nodes in the graph is $2n + 1$. This family of technology diffusion problems (which again is implicitly parameterized by n) is shown in Figure 5.8 and defined as follows:

- The node set is $\{v_1, v_2, \dots, v_{2n+1}\}$.
- The edge set is constructed as follows,
 - The subsets $\{v_1, \dots, v_n\}$ and $\{v_{n+1}, \dots, v_{2n}\}$ form two cliques.
 - Node v_{2n+1} is connected to all other nodes in the graph, *i.e.* edges are $\{v_1, v_{2n+1}\}, \dots, \{v_{2n}, v_{2n+1}\}$.
- The threshold function is
 - for $i \leq 2n$, $\theta(v_i) = n + 2$.
 - $\theta(v_{2n+1}) = 2n + 1$.

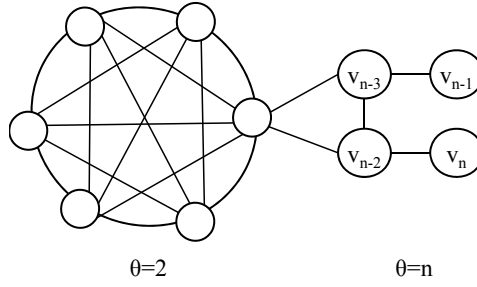


Figure 5.9: Another instance of the technology diffusion problem.

To show this problem is non-submodular, we shall find two disjoint sets S_1 and S_2 such that

$$f(S_1) + f(S_2) < f(S_1 \cup S_2) \quad (5.10)$$

We chose $S_1 = \{v_1, \dots, v_n\}$ and $S_2 = \{v_{2n+1}\}$. Note that S_1 and S_2 are connected, and that $f(S_1) = n$, $f(S_2) = 1$, while $f(S_1 \cup S_2) = 2n + 1$ so that (5.10) holds. \square

The influence function is not supermodular.

Let n be a sufficiently large integer that represents the number of nodes in the graph. Our family of technology diffusion problems G, θ (implicitly parameterized by n) shown in Figure 5.9 and defined as follows:

- The node set is $\{v_1, \dots, v_n\}$.
- The edge set is defined as follows:
 - For any $1 \leq i < j \leq n - 4$, $\{v_i, v_j\}$ is in the edge set, *i.e.* the subgraph induced by $\{v_1, \dots, v_{n-4}\}$ is a complete graph.
 - The remaining edges are $\{v_1, v_{n-3}\}$, $\{v_1, v_{n-2}\}$, $\{v_{n-3}, v_{n-1}\}$, $\{v_{n-2}, v_n\}$, and $\{v_{n-3}, v_{n-2}\}$.

- The threshold function is
 - For $i \leq n - 4$, $\theta(v_i) = 2$.
 - For $i > n - 4$, $\theta(v_i) = n$.

To show this problem is not supermodular, we choose two disjoint sets S_1 and S_2 such that

$$f(S_1) + f(S_2) > f(S_1 \cup S_2) \quad (5.11)$$

We choose $S_1 = \{v_{n-3}\}$ and $S_2 = \{v_{n-2}\}$. Note that S_1 and S_2 are connected, and $f(S_1) = f(S_2) = n - 3$, while $f(S_1 \cup S_2) = n - 2$ so that (5.11) indeed holds. \square

5.6.2 Randomized threshold case

We now consider a modified version of our problem, where, as in [69], we assume that thresholds are chosen uniformly at random:

Definition 5.6.1 (Randomized technology diffusion optimization problem.). *The randomized technology diffusion model is as before, with the exception that nodes choose their thresholds uniformly and independently at random from the set $\{2, 3, \dots, n\}$. Thus, the randomized technology diffusion optimization problem is to find the smallest feasible seedset S in expectation over the choice of thresholds, when G is given as input.*

We follow [69] and let the influence function $f_G(S)$ be the *expected* number of nodes that are eventually activated, *i.e.* $f_G(S) = \mathbb{E}_\theta[f_{G,\theta}(S)]$, where $f_{G,\theta}(S)$ is the number of activated nodes, and expectation is taken over the choice of thresholds. We present two families of problem instances: each family will be on a graph of

diameter at most 4, and will consider connected seedsets. The first family will *fail* to exhibit submodularity of $f_G(S)$, while the second will *fail* to exhibit supermodularity.

The influence function is not submodular.

Let n be a sufficiently large integer such that the number of nodes in the network is $2n + 1$. Our family of G (parameterized by n) is defined as

- The node set is $\{v_1, v_2, \dots, v_{2n+1}\}$.
- The edge set is constructed as follows,
 - The subsets $\{v_1, \dots, v_n\}$ and $\{v_{n+1}, \dots, v_{2n}\}$ form two cliques.
 - The remaining edges are $\{v_{2n+1}, v_1\}$ and $\{v_{2n+1}, v_{2n}\}$.

Notice that this family of graphs is *almost* identical to the non-submodular example presented in the previous section, shown in Figure 5.8, except that now, the middle node v_{2n+1} is only connected to v_1 and v_{2n} . We shall find two disjoint set S_1 and S_2 such that

$$f_G(S_1) + f_G(S_2) < f_G(S_1 \cup S_2). \quad (5.12)$$

Our choice of S_1 and S_2 is $S_1 = \{v_1, \dots, v_n\}$ and $S_2 = \{v_{2n+1}\}$. We start with computing $f_G(S_1)$:

$$\begin{aligned} & f_G(S_1) \\ = & \mathbb{E}[f_{G,\theta}(S_1) \mid \theta(v_{2n+1}) \leq n + 1] \Pr[\theta(v_{2n+1}) \leq n + 1] \\ & + \mathbb{E}[f_{G,\theta}(S_1) \mid \theta(v_{2n+1}) > n + 1] \Pr[\theta(v_{2n+1}) > n + 1] \end{aligned} \quad (5.13)$$

Notice that

$$\mathbb{E}[f_{G,\theta}(S_1) \mid \theta(v_{2n+1}) \leq n+1] = \mathbb{E}[f_{G,\theta}(S_1 \cup S_2)] = f_G(S_1 \cup S_2) \quad (5.14)$$

$$\mathbb{E}[f_{G,\theta}(S_1) \mid \theta(v_{2n+1}) > n+1] = n$$

Therefore, we may rewrite (5.13) as

$$f_G(S_1) = f_G(S_1 \cup S_2) \Pr[\theta(v_{2n+1}) \leq n+1] + n \Pr[\theta(v_{2n+1}) > n+1] = \frac{f_G(S_1 \cup S_2)}{2} + \frac{n}{2}. \quad (5.15)$$

We next move to compute $f_G(S_2)$. To understand how the influence of $S_2 = \{v_{2n+1}\}$ spreads, we condition on the thresholds of its neighbors: $\theta(v_1), \theta(v_{2n})$.

$$\begin{aligned} f_G(S_2) &\leq 1 \cdot \Pr[\theta(v_1) > 2 \cap \theta(v_{2n}) > 2] + (2n+1) \cdot \Pr[\theta(v_1) = 2 \cup \theta(v_{2n}) = 2] \\ &= 1(1 - \frac{1}{2n})(1 - \frac{1}{2n}) + (2n+1)(2\frac{1}{2n}(1 - \frac{1}{2n}) + \frac{1}{2n}\frac{1}{2n}) \\ &= 1 + 2n\frac{1}{2n}(2(1 - \frac{1}{2n}) + \frac{1}{2n}) \leq 3 \end{aligned} \quad (5.16)$$

Therefore, from (5.15) and (5.16) we have

$$f_G(S_1) + f_G(S_2) \leq 3 + \frac{1}{2}(f_G(S_1 \cup S_2) + n) \quad (5.17)$$

Recall that our goal is to show that $f_G(S_1) + f_G(S_2) < f_G(S_1 \cup S_2)$. Using (5.17), we now see that it suffices to prove that

$$f_G(S_1 \cup S_2) > n+6$$

We prove this by conditioning on the event that $S_1 \cup S_2$ activates node v_{2n} :

$$\begin{aligned} f_G(S_1 \cup S_2) &= f_G(S_1 \cup S_2 \cup \{v_{2n}\}) \Pr[\theta(v_{2n}) \leq n+2] + (n+1) \Pr[\theta(v_{2n}) > n+2] \\ &\geq (n+2 + \frac{n-1}{2})\frac{n+1}{2n} + (n+1)\frac{n-1}{2n} \\ &= n+1 + \frac{n+1}{4} \end{aligned}$$

where the first inequality follows because the thresholds of half of the nonseed nodes $\{v_{n+1}, \dots, v_{2n-1}\}$ are $\leq n+1$ in expectation. Thus, we indeed have that S_1 and S_2 are connected and $f_G(S_1) + f_G(S_2) < f_G(S_1 \cup S_2)$ when n is sufficiently large. \square

The influence function is not supermodular.

Let n be a sufficiently large integer such that the number of nodes in the network is $2n+1$. Our family of graphs (parameterized by n) is defined as follows,

- The node set is $\{v_1, v_2, \dots, v_{2n+1}\}$.
- The edge set is constructed as follows,
 - The subsets $\{v_1, \dots, v_n\}$ and $\{v_{n+1}, \dots, v_{2n}\}$ form two cliques.
 - Node v_{2n+1} is connected to all other nodes in the graph.
 - There is an additional edge $\{v_1, v_{2n}\}$.

Notice that this family of graphs is almost identical to the one shown in Figure 5.8, except for the addition of a single edge $\{v_1, v_{2n}\}$. We shall find two disjoint set S_1 and S_2 such that

$$f_G(S_1) + f_G(S_2) > f_G(S_1 \cup S_2). \quad (5.18)$$

Our choice of S_1 and S_2 is $S_1 = \{v_1, \dots, v_n\}$ and $S_2 = \{v_{n+1}, \dots, v_{2n}\}$. Notice that these sets are connected by the edge $\{v_1, v_{2n}\}$. By symmetry we have that $f(S_1) = f(S_2)$, so we start by computing $f_G(S_1)$. Let T be the number of active nodes in S_2 , and let

A be the event that node v_{2n+1} is active.

$$\begin{aligned}
\mathbb{E}[f_{G,\theta}(S_1)] &\geq n + (1 + \mathbb{E}[T|A, S_1 \text{ active}]) \Pr[A|S_1 \text{ active}] \\
&\geq n + (1 + n \cdot \frac{n+1}{2n}) \frac{n}{2n} \\
&= n + \frac{1}{2}(1 + \frac{n+1}{4})
\end{aligned} \tag{5.19}$$

where the second inequality follows because we used the trivial bound $\mathbb{E}[T|A, S_1 \text{ active}] \geq n \frac{n+1}{2n}$ where we ignore all cascading effects; we simply assume that each of the n nodes in S_2 is connected to an active component of size $n + 1$. On the other hand,

$$\mathbb{E}[f_{G,\theta}(S_1 \cup S_2)] \leq 2n + 1 \tag{5.20}$$

Thus we indeed have $f_G(S_1) + f(S_2) \geq 2n + 1 + \frac{n+1}{4} > 2n + 1 = f_G(S_1 \cup S_2)$ for all n . □

5.7 Experiments with the IP of Figure 5.1

Given the prevalence of heuristics like “choose the high degree nodes” in the literature on technology diffusion in communication networks (*e.g.* [10, 19, 47]), we sanity-check our approach against several heuristics. Our goal in the following is to *give evidence* that we can find solutions that are substantially different from known heuristics, and to suggest that our IP could be a promising starting point for the design of new heuristics.

We considered problem instances where (a) $G(V, E)$ is 200-node preferential attachment graph with node outdegree randomly chosen from $\{1, 2, 3, 4\}$ [1], and (b) thresholds θ randomly chosen from

$\{\max\{2, c\}, 2c, 3c, \dots, \lceil \frac{200}{c} \rceil \cdot c\}$. We ran four groups of experiments with threshold step-length parameter c fixed to 1, 5, 10, and 20 respectively. For each group, we used a fresh random preferential attachment graph, and repeated the experiment five times with a fresh random instance of the threshold functions. We solved each of these 20 problem instances using the simple IP formulation presented in Figure 5.1 (with the extra restriction that the highest degree node must be part of the seedset) and the Gurobi IP solver. We compared the result against five natural heuristics that iteratively pick a node u with property X from the set of inactive nodes, add u to the seedset S' , activate u , let u activate as many nodes as possible, and repeats until all nodes are active. We instantiate property X as:

- (a) *degree*: highest degree,
- (b) *degree-threshold*: highest $(\text{degree}) \times (\text{threshold})$,
- (c) *betweenness*: highest betweenness centrality,
- (d) *degree discounted*: highest degree in the subgraph induced by the inactive nodes [22],
- (e) *degree connected*: highest degree and connected to the active nodes.

For each group, Table 5.4 presents the average seedset size and the average Jaccard index $\frac{|S \cap S'|}{|S \cup S'|}$ between IP seedset S and the heuristic seedset S' . We also compute the fraction of nodes in S that are also part of the top- $|S|$ nodes in terms of (a) degree (the row denoted “degree overlap”), and (b) betweenness centrality (“betweenness overlap”). The results of Table 5.4 do indeed give evidence that our IP can return seedsets that are substantially different (and often better), than the seedsets found via heuristics.

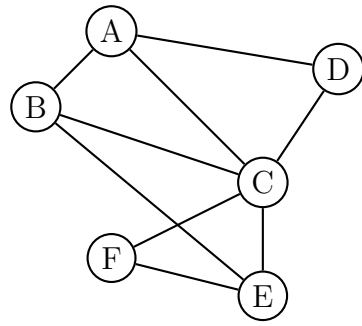
Table 5.4: Comparison of the IP of Figure 5.1 to several heuristics.

threshold step length:	$c = 1$		$c = 5$		$c = 10$		$c = 20$	
	Size	Jaccard	Size	Jaccard	Size	Jaccard	Size	Jaccard
degree	11.8	0.42	20.9	0.36	24.45	0.38	41.75	0.46
degree-threshold	8.95	0.41	15.40	0.42	19.00	0.44	33.25	0.55
betweenness	10.50	0.45	19.65	0.39	24.2	0.38	40.85	0.47
degree discounted	11.2	0.39	21.55	0.34	25.35	0.36	41.60	0.45
degree connected	12.9	0.35	22.65	0.29	25.90	0.33	43.25	0.44
ip_solver	6.45	1	11.15	1	13.75	1	23.45	1
degree overlap		0.44		0.39		0.37		0.39
betweenness overlap		0.47		0.39		0.37		0.40

5.8 Expository examples and figures

We now present examples of the constructions we used in Chapter 5.2 and Chapter 5.3. We start with the problem instance $\{G, \theta\}$ in Figure 5.10, and present a feasible connected activation sequence T for this problem instance. This feasible connected activation sequence T uniquely corresponds to the seedset $S = \{A, D\}$, since these are the only nodes that have $T(u) < \theta(u)$.

The flow graph \mathcal{H} used for the relaxed linear program is shown in Figure 5.11. The solid line is the *threshold line*. The (solid and dotted) trajectories represent some paths that can be used to push some amount of flow $f \in [0, 1]$ between the nodes in \mathcal{H} (*i.e.* so that the flow constraints are satisfied). Notice that every trajectory in \mathcal{H} corresponds to an edge in the original graph G . Let us consider the $(E, 6)$ -flow problem. The solid trajectories in Figure 5.11 illustrate a feasible flow to solve the problem, which we use as the representative flow $\mathcal{F}_{E,6}$. Notice that $(E, 6)$ -flow has demand from two nodes $(E, 5)$ and $(E, 6)$ and thus $\mathcal{F}_{E,6}$ has two sinks. We decompose $\mathcal{F}_{E,6}$ into two paths $\mathcal{P}_1 = (A, 1), (C, 3), (F, 4), (E, 5)$ and $\mathcal{P}_2 = (A, 1), (B, 5), (E, 6)$. The *border node* for path \mathcal{P}_1 is $\text{border}(\mathcal{P}) = (F, 4)$ and the border node for path \mathcal{P}_2 is $\text{border}(\mathcal{P}) = (A, 1)$. Thus, $\beta(E, 6) = \{(A, 1), (F, 4)\}$ and $B(E, 6) = \{A, F\}$.



Threshold function

$$\theta(A) = 5$$

$$\theta(B) = 2$$

$$\theta(C) = 3$$

$$\theta(D) = 5$$

$$\theta(E) = 4$$

$$\theta(F) = 6$$

A connected activation sequence

$$x_{A,1} = 1, (\forall t \neq 1, x_{A,t} = 0)$$

$$x_{B,2} = 1, (\forall t \neq 2, x_{B,t} = 0)$$

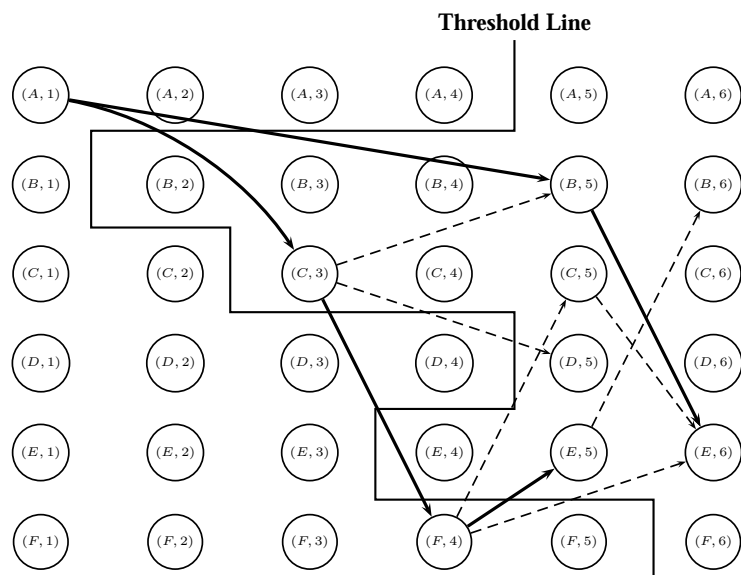
$$x_{C,3} = 1, (\forall t \neq 3, x_{C,t} = 0)$$

$$x_{D,5} = 1, (\forall t \neq 5, x_{D,t} = 0)$$

$$x_{E,6} = 1, (\forall t \neq 6, x_{E,t} = 0)$$

$$x_{F,4} = 1, (\forall t \neq 4, x_{F,t} = 0)$$

Figure 5.10: A problem instance and a feasible connected activation sequence.

Figure 5.11: The \mathcal{H} graph and the trajectories of flows.

Bibliography

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74, Jan 2002.
- [2] D. Aldous. Some inequalities for reversible markov chains. *Journal of London Mathematical Society*, 25:564–576, 1982.
- [3] Noga Alon, Dana Moshkovitz, and Shmuel Safra. Algorithmic construction of sets for k-restrictions. *ACM Trans. Algorithms*, 2:153–177, April 2006.
- [4] O. S. M. Alves, F. P. Machado, S. Yu. Popov, and K. Ravishankar. The shape theorem for the frog model with random initial configuration. *arXiv:math/0110280v1*, October 2001.
- [5] O. S. M. Alves, S. Yu. Popov, and F. P. Machado. The shape theorem for the frog model. *Annals of Applied Probability*, 12(2):533–546, 2002.
- [6] C. Arackaparambil, J. Brody, and A. Chakrabarti. Functional monitoring without monotonicity. In *Proc. of ICALP*, 2009.
- [7] Tamio Arai, Eiichi Yoshida, and Jun Ota. Information diffusion by local communication of multiple mobile robots. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 1993*, pages 535–540, 1993.
- [8] Brice Augustin, Balachander Krishnamurthy, and Walter Willinger. IXPs: Mapped? In *IMC*, 2009.
- [9] K. Avrachenkov, B. Ribeiro, and D. Towsley. Improving random walk estimation accuracy with uniform restart. In *7th Workshop on Algorithms and Models for the Web Graphs (WAW 2010)*, 2010.
- [10] Ioannis Avramopoulos, Martin Suchara, and Jennifer Rexford. How small groups can secure interdomain routing. Technical report, Princeton University Comp. Sci., 2007.
- [11] N. Bansal, A. Blum, S. Chawla, and A. Meyerson. Approximation algorithms for deadline-tsp and vehicle routing with time-windows. In *Proceedings of the*

- thirty-sixth annual ACM symposium on Theory of computing*, pages 166–174. ACM, 2004.
- [12] Boaz Barak, Sharon Goldberg, and David Xiao. Protocols and lower bounds for failure localization in the Internet. In *IACR EUROCRYPT*, 2008.
- [13] F.M. Bass. A new product growth model for consumer durables. *Management Science*, pages 215–27, 1969.
- [14] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [15] A. Blum, P. Chalasani, D. Coppersmith, W. Pulleyblank, P. Raghavan, and M Sudan. The minimum latency problem. In *SDM*, 1994.
- [16] Béla Bollobás and Imre Leader. Compressions and isoperimetric inequalities. *Journal of Combinatorial Theory, Series A*, 56:47–62, 1991.
- [17] Randal E. Bryant, Randy H. Katz, and Edward D. Lazowska. Big-data computing: creating revolutionary breakthroughs in commerce, science, and society. *CRA Report*, 2008.
- [18] K Butler, T Farley, P McDaniel, and J Rexford. A survey of BGP security issues and solutions. *Proceedings of the IEEE*, 2010.
- [19] Haowen Chang, Debabrata Dash, Adrian Perrig, and Hui Zhang. Modeling adoptability of secure BGP protocol. In *Sigcomm*, 2006.
- [20] N. Chen. On the approximability of influence social networks. In *ACM-SIAM Symposium on Discrete Algorithms*, 2008.
- [21] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincón, X. Sun, Y. Wang, Wei. W., and Yuan. Y. Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, 2011.
- [22] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proc. 15th Conference on Knowledge discovery and data mining*, KDD '09, pages 199–208. ACM, 2009.
- [23] Ying-Ju Chi, Ricardo Oliveira, and Lixia Zhang. Cyclops: The Internet AS-level observatory. *ACM SIGCOMM CCR*, 2008.
- [24] David D. Clark, John Wroclawski, Karen R. Sollins, and Robert Braden. Tussle in cyberspace: defining tomorrow’s Internet. *Trans. on Networking*, 2005.
- [25] Andrea E. F. Clementi, Angelo Monti, and Riccardo Silvestri. Fast flooding over manhattan. In *PODC*, pages 375–383. ACM, 2010.

- [26] Andrea E. F. Clementi, Francesco Pasquale, and Riccardo Silvestri. Manets: High mobility can make up for low transmission power. In *ICALP (2)*, volume 5556 of *Lecture Notes in Computer Science*, pages 387–398. Springer, 2009.
- [27] G. Cormode and M. Garofalakis. Sketching streams through the net: Distributed approximate query tracking. In *Proc. of International Conference on Very Large Databases*, 2005.
- [28] G. Cormode, M. Garofalakis, S. Muthukrishnan, and R. Rastogi. Holistic aggregates in a networked world: Distributed tracking of approximate quantiles. In *Proc. of SIGMOD*, June 2005.
- [29] G. Cormode, S. Muthukrishnan, and W. Zhuang. What’s different: Distributed, continuous monitoring of duplicate-resilient aggregates on data streams. In *Proc. IEEE International Conference on Data Engineering*, 2006.
- [30] G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed functional monitoring. In *Proc. of SODA*, 2008.
- [31] G. Cormode, S. Muthukrishnan, K. Yi, and Q. Zhang. Optimal sampling from distributed streams. In *Proc. of PODS*, June 2010.
- [32] Graham Cormode, Minos Garofalakis, Peter J. Haas, and Chris Jermaine. Synopses for massive data: Samples, histograms, wavelets, sketches. *Foundations and Trends in Databases*, 2011.
- [33] S. Deering and R. Hinden. RFC 2460: Internet Protocol, Version 6 (IPv6) Specification. <http://www.ietf.org/rfc/rfc2460.txt>, 1998.
- [34] J. W. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *NUMERISCHE MATHEMATIK*, 51(3):251–289, 1987.
- [35] M. Desrochers, J.K. Lenstra, M.W.P. Savelsbergh, and F. Soumis. Vehicle routing with time windows: Optimization and approximation. Technical report, Department of Operations Research and System Theory, Centrum voor Wiskunde en Informatica, 1987.
- [36] Xenofontas Dimitropoulos, Dmitri Krioukov, Marina Fomenkov, Bradly Huffaker, Young Hyun, and kc claffy. AS relationships: Inference and validation. *ACM SIGCOMM Computer Communication Review*, JAN 2007.
- [37] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proc. 7th Conf on Knowledge discovery and data mining*, KDD ’01, pages 57–66, New York, NY, USA, 2001. ACM.

- [38] Benjamin Edelman. Running out of numbers: Scarcity of ip addresses and what to do about it. Technical report, Harvard Business School, 2009.
- [39] H. A. Elmore, L. J. Camp, and B. P. Stephens. Diffusion and adoption of ipv6 in the arin region. In *Workshop on the Economics of Internet Security*, 2008.
- [40] J. Farrell and G. Saloner. Standardization, compatibility, and innovation. *The RAND Journal of Economics*, pages 70–83, 1985.
- [41] FCC. The communications security, reliability and interoperability council iii working group 6: Secure bgp deployment. Technical report, March 2012.
- [42] William Feller. *An introduction to probability theory and its applications, Volume 1*. Wiley series in probability and mathematical statistics. John Wiley & sons, 1951.
- [43] A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *Annals of Applied Probability*, 1, Number 1:62–87, 1991.
- [44] Greg N. Frederickson and Barry Wittman. Approximation algorithms for the traveling repairman and speeding deliveryman problems. *Algorithmica*, 62(3-4):1198–1221, 2012.
- [45] A. Frieze, P. Melsted, and M. Mitzenmacher. An analysis of random-walk cuckoo hashing. *SIAM Journal of Computing*, 2011.
- [46] P. Gill, M. Schapira, and S. Goldberg. Modeling on quicksand: dealing with the scarcity of ground truth in interdomain routing data. *ACM SIGCOMM Computer Communication Review*, 42(1):40–46, 2012.
- [47] Phillipa Gill, Michael Schapira, and Sharon Goldberg. Let the market drive deployment: A strategy for transistioning to BGP security. *SIGCOMM’11*, 2011.
- [48] D. Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4), 1997.
- [49] O. Goldreich and D. Ron. On testing expansion in bounded degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 2000.
- [50] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, May 1978.
- [51] M. B. Greenwald and S. Khanna. Space-efficient online computation of quantile summaries. In *Proc. of SIGMOD*, pages 58–66, 2001.

- [52] M. B. Greenwald and S. Khanna. Power-conserving computation of order-statistics over sensor networks. In *Proc. of PODS*, 2004.
- [53] Roch Guérin and Kartik Hosanagar. Fostering ipv6 migration through network quality differentials. *SIGCOMM Comput. Commun. Rev.*, 40:17–25, June 2010.
- [54] J. Hao and J B. Orlin. A faster algorithm for finding the minimum cut in a graph. In *SODA*, 1992.
- [55] A. Healy. Randomness efficient sampling within nc^1 . *Computational Complexity*, 17(1), 2008.
- [56] W. Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, pages 13–30, March 1963.
- [57] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [58] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2001.
- [59] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.
- [60] M.P. Howarth, P. Flegkas, G. Pavlou, N. Wang, P. Trimintzios, D. Griffin, J. Griem, M. Boucadair, P. Morand, H. Asgari, and P. Georgatsos. Provisioning for inter-domain quality of service: the MESCAL approach. *IEEE Communications Magazine*, June 2005.
- [61] Z. Huang, K. Yi, and Q. Zhang. Randomized algorithms for tracking distributed count, frequencies, and ranks. In *arXiv:1108.3413v1*, Aug 2011.
- [62] Barry D. Hughes. *Random Walks and Random Environments*. Volume 1: Random Walks. Oxford University Press, 1995.
- [63] G Huston. Stacking it up: Experimental observations on the operation of dual stack services. In *NANOG’52*, 2011.
- [64] M. Jerrum and A. Sinclair. *The Markov chain Monte Carlo method: an approach to approximate counting and integration*. PWS Publishing, Boston, MA, USA, 1996.
- [65] Y. Jin, S. Sen, R. Guerin, K. Hosanagar, and Z-L. Zhang. Dynamics of competition between incumbent and emerging network technologies. *NetEcon*, 2008.

-
- [66] D. Joseph, N. Shetty, J. Chuang, and I. Stoica. Modeling the adoption of new network architectures. In *CoNEXT'07: Conference on emerging Networking EXperiments and Technologies*, 2007.
- [67] N. Kahale. Large deviation bounds for markov chains. *Combinatorics, Probability, and Computing*, 6(4), 1997.
- [68] M.L. Katz and C. Shapiro. Network externalities, competition, and compatibility. *The American economic review*, 75(3):424–440, 1985.
- [69] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *ACM SIGKDD*, 2003.
- [70] S. Kent, C. Lynn, and K. Seo. Secure border gateway protocol (S-BGP). *JSAC*, 2000.
- [71] Harry Kesten and Vladas Sidoravicius. The spread of a rumor or infection in a moving population. *Annals of Probability*, 33(6):pp. 2402–2462, 2005.
- [72] Abdelmajid Khelil, Christian Becker, Jing Tian, and Kurt Rothermel. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, MSWiM, pages 54–60. ACM, 2002.
- [73] T. Konstantopoulos. *Markov Chains and Random Walks*. Lecture notes, 2009.
- [74] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.
- [75] C. A. Leon and F. Perron. Optimal hoeffding bounds for discrete reversible markov chains. *Annals of Applied Probability*, 14(2), 2004.
- [76] M. Lepinski, editor. *BGPSEC Protocol Specification*. IETF Network Working Group, Internet-Draft, March 2011. Available from <http://tools.ietf.org/html/draft-lepinski-bgpsec-protocol-00>.
- [77] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2005.
- [78] D. Leventhal and A. S. Lewis. Randomized methods for linear constraints: Convergence rates and conditioning. *Mathematics of Operations Research*, 35(3):641–654, 2010.
- [79] P. Lezaud. Chernoff-type bound for finite markov chains. *Annals of Applied Probability*, 8(3):849–867, 1998.

- [80] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, and J. Hellerstein. Graphlab: A new framework for parallel machine learning. In *Proc. of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2010.
- [81] Michael W. Mahoney. Algorithmic and statistical perspectives on large-scale data analysis. *arXiv:1010.1609v1*, 2010.
- [82] Carolyn Duffy Marsan. White house issues ipv6 directive. *Network World*, September 28, 2010.
- [83] B Metcalfe. Metcalfe’s law: A network becomes more valuable as it reaches more users. *InfoWorld*, 1995.
- [84] Michael Mitzenmacher and Eli Upfal. *Probability and computing: randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.
- [85] A Mohaisen, A. Yun, and Y Kim. Measuring the mixing time of social graphs. In *IMC’10 Proceedings of the 10th Annual Conference on Internet Measurement*, 2010.
- [86] Stephen Morris. Contagion. *The Review of Economic Studies*, Jan., 2000.
- [87] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Computer Science*, 2005.
- [88] C. Olston, J. Jiang, and J. Widom. Adaptive filters for continuous queries over distributed data streams. In *Proc. ACM SIGMOD International Conference on Management of Data*, 2003.
- [89] Andy Ozment and Stuart E. Schechter. Bootstrapping the adoption of internet security protocols. In *The Fifth Workshop on the Economics of Information Security (WEIS 2006)*., 2006.
- [90] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
- [91] Yuval Peres, Alistair Sinclair, Perla Sousi, and Alexandre Stauffer:. Mobile geometric graphs: Detection, coverage and percolation. In *SODA*, pages 412–428. SIAM, 2011.
- [92] Alberto Pettarin, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Tight bounds on information dissemination in sparse mobile networks. *arXiv:1101.4609v2*, 2011.

-
- [93] Manuel Gomez Rodriguez and Bernhard Schölkopf. Influence maximization in continuous time diffusion networks. In *29th International Conference on Machine Learning (ICML)*, 2012.
 - [94] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 lessons from 10 years of measuring and modeling the internet’s autonomous systems. *Selected Areas in Communications, IEEE Journal on*, 29(9):1810–1821, 2011.
 - [95] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes*. Chapman & Hall, 1994.
 - [96] Purnamrita Sarkar and Andrew W. Moore. Dynamic social network analysis using latent space models. *SIGKDD Explorations*, 7(2):31–40, 2005.
 - [97] T C Schelling. *Micromotives and Macrobehavior*. Norton, 1978.
 - [98] R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *Ann. Statist*, 2(1):39–48, 1974.
 - [99] Y. Shavitt and E. Shir. Dimes: Let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35(5):71–74, 2005.
 - [100] Alistair Sinclair. Improved bounds for mixing rates of marked chains and multi-commodity flow. In *LATIN*, volume 583 of *Lecture Notes in Computer Science*, pages 474–487. Springer, 1992.
 - [101] Alistair Sinclair and Alexandre Stauffer. Mobile geometric graphs, and detection and communication problems in mobile wireless networks. *arXiv:1005.1117v2*, 2010.
 - [102] Twitter Search Team. The engineering behind twitter’s new search experience, 2011.
 - [103] C. Tekin and M. Liu. Online algorithms for the multi-armed bandit problem with markovian rewards. In *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010.
 - [104] S. Trithapura and D. P. Woodruff. Optimal random sampling from distributed streams revisited. In *Proc. of DISC*, Roma, Italy, Sep 2011.
 - [105] Salil P. Vadhan. Pseudorandomness, available at <http://people.seas.harvard.edu/~salil/pseudorandomness/>. 2010.
 - [106] R. Wagner. Tail estimates for sums of variables sampled by a random walk. *Combinatorics, Probability, and Computing*, 17(2), 2008.

-
- [107] Yi Wang, Shyam Kapadia, and Bhaskar Krishnamachari. Infection spread in wireless networks with random and adversarial node mobilities. In *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, MobilityModels '08, pages 17–24. ACM, 2008.
 - [108] A. Wigderson and D. Xiao. A randomness-efficient sampler for matrix valued functions and applications. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005.
 - [109] David Williams. *Probability with martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.
 - [110] David P. Williamson and D. B. Shmoys. *The design of approximation algorithms*. Cambridge University Press, 2010.
 - [111] A Yaar, A Perrig, and D Song. SIFF: a stateless internet flow filter to mitigate ddos flooding attacks. *IEEE Symposium on Security and Privacy*, 2004.
 - [112] K. Yi and Q. Zhang. Optimal tracking of distributed heavy hitters and quantiles. In *Proc. of PODS*, June 2009.

Appendix A

Probability review

A.1 Concentration bounds

Below we describe several concentration bounds we used in the thesis. These results could be found in standard textbooks such as [42, 84].

Theorem A.1.1 (Chernoff bounds). *Let X_1, \dots, X_n be independent Poisson trials with $\Pr[X_i] = p_i$. Let $X = \sum_{i \leq n} X_i$ and $\mu = \mathbb{E}[X]$. Then the following Chernoff bounds hold:*

- For $0 < \delta < 1$,

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2 \exp(-\mu\delta^2/3).$$

- For $R \geq 6\mu$,

$$\Pr[X \geq R] \leq 2^{-R}.$$

Theorem A.1.2 (Chernoff bounds for dependent variables). *Let X_1, \dots, X_n be possibly dependent Poisson trials with $\Pr[X_i = 1 \mid X_1, \dots, X_{i-1}] \geq p$. Let $X = \sum_{i \leq n} X_i$*

and $\mu = np$. Then the following Chernoff bound holds:

- For $0 < \delta < 1$,

$$\Pr[X \leq (1 - \delta)\mu] \leq \exp(-\mu\delta^2/2).$$

On the other hand, if $\Pr[X_i = 1 \mid X_1, \dots, X_{i-1}] \leq p$, the following bound holds:

- For any $\delta > 0$,

$$\Pr[X > (1 + \delta)\mu] \leq \exp(-\mu\delta^2/4).$$

Hoeffding's Inequality

Let $-\infty < a_i \leq b_i < \infty$ be a given sequence of real numbers for $1 \leq i \leq t$. The following inequality was established by Hoeffding [56].

Theorem A.1.3. Suppose X_1, X_2, \dots, X_t is a sequence of independent random variables such that $a_i \leq X_i \leq b_i$ for every $1 \leq i \leq t$. Let $\mu = \mathbb{E}[\sum_{i \leq t} X_i]$. Then, the following inequality holds, for every $x \geq 0$,

$$\Pr\left[\sum_{i \leq t} X_i - \mu \geq x\right] \leq \exp\left(-\frac{2x^2}{\sum_{i \leq t} (b_i - a_i)^2}\right).$$

The following special case is also used in the thesis.

Corollary A.1.4. Assume that in addition $a_i = -1$ and $b_i = 1$, for every $1 \leq i \leq t$.

Then,

$$\Pr\left[\sum_{i \leq t} X_i - \mu \geq x\right] \leq \exp\left(-\frac{x^2}{2t}\right), \text{ for every } x \geq 0.$$

Sampling Without Replacement

Similar type of bounds as in earlier section hold also for sampling without replacement. In particular, the following is a result of Hoeffding [56] (Section 6) for sampling from a finite population.

Theorem A.1.5. *Suppose X_1, X_2, \dots, X_t is a sample drawn by uniform random sampling without replacement from a set of values $\{c_1, c_2, \dots, c_n\}$ such that for some $-\infty < a < b < \infty$, it holds $a \leq c_i \leq b$, for every $1 \leq i \leq n$. Then, the bound of Theorem A.1.3 holds with $\mu = (\frac{1}{n} \sum_{i \leq n} c_i)t$, i.e. for every $x \geq 0$,*

$$\Pr[\sum_{i \leq t} X_i - \mu \geq x] \leq \exp(-\frac{2x^2}{(b-a)^2 t}).$$